PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/54, 9/10 | A1

(11) International Publication Number: WO 98/00549

(43) International Publication Date: 8 January 1998 (08.01.98)

(21) International Application Number: PCT/AU97/00402

(22) International Filing Date: 24 June 1997 (24.06.97)

(30) Priority Data:
PO 0699 27 June 1996 (27.06.96) AU

(71) Applicants *(for all designated States except US)*: THE AUSTRALIAN NATIONAL UNIVERSITY [AU/AU]; Acton, ACT 2601 (AU). COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION [AU/AU]; Limestone Avenue, Campbell, ACT 2601 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ARIOLI, Antonio [AU/AU]; 8 Burrinjuck Crescent, Duffy, ACT 2611 (AU). WILLIAMSON, Richard, Edward [AU/AU]; "Wybalena", Butts Road, Murrumbateman, NSW 2582 (AU). BETZNER, Andreas, Stefan [DE/AU]; 40 Dallachy Place, Page, ACT 2614 (AU). PENG, Liangcai [AU/AU]; 3/12 Masson Street, Turner, ACT 2612 (AU).

(74) Agents: SLATTERY, John, Michael et al.; Davies Collison Cave, 1 Little Collins Street, Melbourne, VIC 3000 (AU).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MANIPULATION OF CELLULOSE AND/OR β-1,4-GLUCAN

(57) Abstract

The present invention relates generally to isolated genes which encode polypeptides involved in cellulose biosynthesis in plants and transgenic plants expressing same in sense or antisense orientation, or as ribozymes, co-suppression or gene-targeting molecules. More particularly, the present invention is directed to a nucleic acid molecule isolated from *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.* which encode an enzyme which is important in cellulose biosynthesis, in particular the cellulose synthase enzyme and homologues, analogues and derivatives thereof and uses of same in the production of transgenic plants expressing altered cellulose biosynthetic properties.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

"MANIPULATION OF CELLULOSE AND/OR β-1,4-GLUCAN"

The present invention relates generally to isolated genes which encode polypeptides involved in cellulose biosynthesis and transgenic organisms expressing same in sense or antisense orientation, or as ribozymes, co-suppression or gene-targeting molecules. More particularly, the present invention is directed to a nucleic acid molecule isolated from *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.* which encode an enzyme which is important in cellulose biosynthesis, in particular the cellulose synthase enzyme and homologues, analogues and derivatives thereof and uses of same in the production of transgenic plants expressing altered cellulose biosynthetic properties.

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description. Sequence identity numbers (SEQ ID Nos.) for the nucleotide and amino acid sequences referred to in the specification are defined after the bibliography.

Throughout the specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising" will be understood to imply the inclusion of a stated element or integer or group of elements or integers but not the exclusion of any other element or integer or group of elements or integers.

Cellulose, the world's most abundant biopolymer, is the most characteristic component of plant cell walls in so far as it forms much of the structural framework of the cell wall. Cellulose is comprised of crystalline β-1,4-glucan microfibrils. The crystalline microfibrils are extremely strong and resist enzymic and mechanical degradation, an important factor in determining the nutritional quantity, digestibility and palatability of animal and human foodstuffs. As cellulose is also the dominant structural component of industrially-important plant fibres, such as cotton, flax, hemp, jute and the timber crops such as *Eucalyptus ssp.* and *Pinus ssp.*, amongst others, there is considerable economic benefit to be derived from the

manipulation of cellulose content and/or quantity in plants. In particular, the production of food and fibre crops with altered cellulose content are highly desirable objectives.

The synthesis of cellulose involves the β-1,4-linkage of glucose monomers, in the form of a nucleoside diphospoglucose such as UDP-glucose, to a pre-existing cellulose chain, catalysed by the enzyme cellulose synthase.

Several attempts to identify the components of the functional cellulose synthase in plants have failed, because levels of β-1,4-glucan or crystalline cellulose produced in such assays have hitherto been too low to permit enzyme purification for protein sequence determination. Insufficient homology between bacterial β-1,4-glucan synthase genes and plant cellulose synthase genes has also prevented the use of hybridisation as an approach to isolating the plant homologues of bacterial β-1,4-glucan (cellulose) synthases.

Furthermore, it has not been possible to demonstrate that the cellulose synthase enzyme from plants is the same as, or functionally related to, other purified and characterised enzymes involved in polysaccharide biosynthesis. As a consequence, the cellulose synthase enzyme has not been isolated from plants and, until the present invention, no nucleic acid molecule has been characterised which functionally-encodes a plant cellulose synthase enzyme.

In work leading up to the present invention, the inventors have generated several novel mutant *Arabidopsis thaliana* plants which are defective in cellulose biosynthesis. The inventors have further isolated a cellulose synthase gene designated *RSW1*, which is involved in cellulose biosynthesis in *Arabidopsis thaliana*, and homologous sequences in *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp*. The isolated nucleic acid molecules of the present invention provide the means by which cellulose content and structure may be modified in plants to produce a range of useful fibres suitable for specific industrial purposes, for example increased decay resistance of timber and altered digestibility of foodstuffs, amongst others.

Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which encodes, or is complementary to a sequence which encodes a polypeptide of the cellulose biosynthetic pathway or a functional homologue, analogue or derivative thereof.

The nucleic acid molecule of the invention may be derived from a prokaryotic source or a eukaryotic source.

Those skilled in the art will be aware that cellulose production requires not only the presence of a catalytic subunit, but also its activation and organisation into arrays which favour the crystallization of glucan chains. This organisation is radically different between bacteria, which possess linear arrays, and higher plants, which possess hexameric clusters or "rosettes", of glucan chains. The correct organisation and activation of the bacterial enzyme may require many factors which are either not known, or alternatively, not known to be present in plant cells, for example specific membrane lipids to impart an active conformation on the enzyme complex or protein, or the bacterial c-di-GMP activation system. Accordingly, the use of a plant-derived sequence in eukaryotic cells such as plants provides significant advantages compared to the use of bacterially-derived sequences.

Accordingly, the present invention does not extend to known genes encoding the catalytic subunit of *Agrobacterium tumefaciens* or *Acetobacter xylinum* or *Acetobacter pasteurianus* cellulose synthase, or the use of such known bacterial genes and polypeptides to manipulate cellulose.

Preferably, the subject nucleic acid molecule is derived from an eukaryotic organism.

In a more preferred embodiment of the invention, the isolated nucleic acid molecule of the invention encodes a plant cellulose synthase or a catalytic subunit thereof, or a homologue, analogue or derivative thereof.

More preferably, the isolated nucleic acid molecule encodes a plant cellulose synthase polypeptide which is associated with the primary cell wall of a plant cell. In an alternative preferred embodiment, the nucleic acid molecule of the invention encodes a plant cellulose synthase or catalytic subunit thereof which is normally associated with the secondary cell wall of a plant cell.

In a more preferred embodiment, the nucleic acid molecule of the invention is a cDNA molecule, genomic clone, mRNA molecule or a synthetic oligonucleotide molecule.

In a particularly preferred embodiment, the present invention provides an isolated nucleic acid molecule which encodes or is complementary to a nucleic acid molecule which encodes the *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.* wheat, barley or maize cellulose synthase enzyme or a catalytic subunit thereof or a polypeptide component, homologue, analogue or derivative thereof.

As exemplified herein, the present inventors have identified cellulose biosynthesis genes in maize, wheat, barley, rice, cotton, *Brassica ssp.* and *Eucalyptus ssp.*, in addition to the specific *Arabidopsis thaliana RSW1* gene sequence which has been shown to be particularly useful for altering cellulose and/or β-1,4-glucan and/or starch levels in cells.

Hereinafter the term "polypeptide of the cellulose biosynthetic pathway" or similar term shall be taken to refer to a polypeptide or a protein or a part, homologue, analogue or derivative thereof which is involved in one or more of the biosynthetic steps leading to the production of cellulose or any related β-1,4-glucan polymer in plants. In the present context, a polypeptide of the cellulose biosynthetic pathway shall also be taken to include both an active enzyme which contributes to the biosynthesis of cellulose or any related β-1,4-glucan polymer in plants and to a polypeptide component of such an enzyme. As used herein, a polypeptide of the cellulose biosynthetic pathway thus includes cellulose synthase. Those skilled in the art will be aware of other cellulose biosynthetic pathway polypeptides in plants.

The term "related β-1,4-glucan polymer" shall be taken to include any carbohydrate molecule comprised of a primary structure of β-1,4-linked glucose monomers similar to the structure of the components of the cellulose microfibril, wherein the relative arrangement or relative configuration of the glucan chains may differ from their relative configuration in microfibrils of cellulose. As used herein, a related β-1,4-glucan polymer includes those β-1,4-glucan polymers wherein individual β-1,4-glucan microfibrils are arranged in an anti-parallel or some other relative configuration not found in a cellulose molecule of plants and those non-crystalline β-1,4-glucans described as lacking the resistance to extraction and degradation that characterise cellulose microfibrils.

The term "cellulose synthase" shall be taken to refer to a polypeptide which is required to catalyse a β-1,4-glucan linkage to a cellulose microfibril.

Reference herein to "gene" is to be taken in its broadest context and includes:

(i) a classical genomic gene consisting of transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (i.e. introns, 5′- and 3′- untranslated sequences); or

(ii) mRNA or cDNA corresponding to the coding regions (i.e. exons) and 5′- and 3′- untranslated sequences of the gene.

The term "gene" is also used to describe synthetic or fusion molecules encoding all or part of a functional product.

In the present context, the term "cellulose gene" or "cellulose genetic sequence" or similar term shall be taken to refer to any gene as hereinbefore defined which encodes a polypeptide of the cellulose biosynthetic pathway and includes a cellulose synthase gene.

The term "cellulose synthase gene" shall be taken to refer to any cellulose gene which specifically encodes a polypeptide which is a component of a functional enzyme having cellulose synthase activity i.e. an enzyme which catalyses a β-1,4-glucan linkage to a

cellulose microfibril.

Preferred cellulose genes may be derived from a naturally-occurring cellulose gene by standard recombinant techniques. Generally, a cellulose gene may be subjected to mutagenesis to produce single or multiple nucleotide substitutions, deletions and/or additions. Nucleotide insertional derivatives of the cellulose synthase gene of the present invention include 5' and 3' terminal fusions as well as intra-sequence insertions of single or multiple nucleotides. Insertional nucleotide sequence variants are those in which one or more nucleotides are introduced into a predetermined site in the nucleotide sequence although random insertion is also possible with suitable screening of the resulting product. Deletional variants are characterised by the removal of one or more nucleotides from the sequence. Substitutional nucleotide variants are those in which at least one nucleotide in the sequence has been removed and a different nucleotide inserted in its place. Such a substitution may be "silent" in that the substitution does not change the amino acid defined by the codon. Alternatively, substituents are designed to alter one amino acid for another similar acting amino acid, or amino acid of like charge, polarity, or hydrophobicity.

As used herein, the term "derived from" shall be taken to indicate that a particular integer or group of integers has originated from the species specified, but has not necessarily been obtained directly from the specified source.

For the present purpose, "homologues" of a nucleotide sequence shall be taken to refer to an isolated nucleic acid molecule which is substantially the same as the nucleic acid molecule of the present invention or its complementary nucleotide sequence, notwithstanding the occurrence within said sequence, of one or more nucleotide substitutions, insertions, deletions, or rearrangements.

"Analogues" of a nucleotide sequence set forth herein shall be taken to refer to an isolated nucleic acid molecule which is substantially the same as a nucleic acid molecule of the present invention or its complementary nucleotide sequence, notwithstanding the occurrence of any

non-nucleotide constituents not normally present in said isolated nucleic acid molecule, for example carbohydrates, radiochemicals including radionucleotides, reporter molecules such as, but not limited to DIG, alkaline phosphatase or horseradish peroxidase, amongst others.

"Derivatives" of a nucleotide sequence set forth herein shall be taken to refer to any isolated nucleic acid molecule which contains significant sequence similarity to said sequence or a part thereof. Generally, the nucleotide sequence of the present invention may be subjected to mutagenesis to produce single or multiple nucleotide substitutions, deletions and/or insertions. Nucleotide insertional derivatives of the nucleotide sequence of the present invention include 5′ and 3′ terminal fusions as well as intra-sequence insertions of single or multiple nucleotides or nucleotide analogues. Insertional nucleotide sequence variants are those in which one or more nucleotides or nucleotide analogues are introduced into a predetermined site in the nucleotide sequence of said sequence, although random insertion is also possible with suitable screening of the resulting product being performed. Deletional variants are characterised by the removal of one or more nucleotides from the nucleotide sequence. Substitutional nucleotide variants are those in which at least one nucleotide in the sequence has been removed and a different nucleotide or nucleotide analogue inserted in its place.

The present invention extends to the isolated nucleic acid molecule when integrated into the genome of a cell as an addition to the endogenous cellular complement of cellulose synthase genes. The said integrated nucleic acid molecule may, or may not, contain promoter sequences to regulate expression of the subject genetic sequence.

The isolated nucleic acid molecule of the present invention may be introduced into and expressed in any cell, for example a plant cell, fungal cell, insect cell, animal cell, yeast cell or bacterial cell. Those skilled in the art will be aware of any moficiations which are required to the codon usage or promoter sequences or other regulatory sequences, in order for expression to occur in such cells.

Another aspect of the present invention is directed to a nucleic acid molecule which comprises

a sequence of nucleotides corresponding or complementary to any one or more of the sequences set forth in SEQ ID Nos:1, 3, 4, 5, 7, 9, 11, or 13, or having at least about 40%, more preferably at least about 55%, still more preferably at least about 65%, yet still more preferably at least about 75-80% and even still more preferably at least about 85-95% nucleotide similarity to all, or a part thereof.

According to this aspect of the invention, said nucleic acid molecule encodes, or is complementary to a nucleotide sequence encoding, a polypeptide of the cellulose biosynthetic pathway in a plant or a homologue, analogue or derivative thereof.

Preferably, a nucleic acid molecule which is at least 40% related to any one or more of the sequences set forth in SEQ ID Nos:1, 3, 4, 5, 7, 9, 11, or 13 comprises a nucleotide sequence which encodes or is complementary to a sequence which encodes a plant cellulose synthase, more preferably a cellulose synthase which is associated with the primary or the secondary plant cell wall of the species from which it has been derived.

Furthermore, the nucleic acid molecule according to this aspect of the invention may be derived from a monocotyledonous or dicotyledonous plant species. In a particularly preferred embodiment, the nucleic acid molecule is derived from *Arabidopsis thaliana, Oryza sativa,* wheat, barley, maize, *Brassica ssp., Gossypium hirsutum* (cotton) or *Eucalyptus ssp.*, amongst others.

For the purposes of nomenclature, the nucleotide sequence shown in SEQ ID NO:1 relates to a cellulose gene as hereinbefore defined which comprises a cDNA sequence designated T20782 and which is derived from *Arabidopsis thaliana*. The amino acid sequence set forth in SEQ ID NO:2 relates to the polypeptide encoded by T20782.

The nucleotide sequence set forth in SEQ ID NO:3 relates to the nucleotide sequence of the complete *Arabidopsis thaliana* genomic gene *RSW1*, including both intron and exon sequences. The nucleotide sequence of SEQ ID NO:3 comprises exons 1-14 of the genomic

gene and includes 2295bp of 5'-untranslated sequences, of which approximately the first 1.9kb comprises *RSW1* promoter sequence (there is a putative TATA box motif at positions 1843-1850 of SEQ ID NO:3). The nucleotide sequence set forth in SEQ ID NO:3 is derived from the cosmid clone 23H12. This sequence is also the genomic gene equivalent of SEQ ID Nos:1 and 5.

The nucleotide sequence set forth in SEQ ID NO:4 relates to the partial nucleotide sequence of a genomic gene variant of *RSW1*, derived from cosmid clone 12C4. The nucleotide sequence of SEQ ID NO:4 comprises exon sequence 1-11 and part of exon 12 of the genomic gene sequence and includes 862bp of 5'-untranslated sequences, of which approximately 700 nucleotides comprise *RSW1* promoter sequences (there is a putative TATA box motif at positions 668-673 of SEQ ID NO:4). The genomic gene sequence set forth in SEQ ID NO:4 is the equivalent of the cDNA sequence set forth in SEQ ID NO:7 (i.e. cDNA clone *Ath*-A).

The nucleotide sequence shown in SEQ ID NO:5 relates to a cellulose gene as hereinbefore defined which comprises a cDNA equivalent of the *Arabidopsis thaliana RSW1* gene set forth in SEQ ID NO:3. The amino acid sequence set forth in SEQ ID NO:6 relates to the polypeptide encoded by the wild-type *RSW1* gene sequences set forth in SEQ ID Nos:3 and 5.

The nucleotide sequence shown in SEQ ID NO:7 relates to a cellulose gene as hereinbefore defined which comprises a cDNA equivalent of the *Arabidopsis thaliana RSW1* gene set forth in SEQ ID NO:4. The nucleotide sequence is a variant of the nucleotide sequences set forth in SEQ ID Nos:3 and 5. The amino acid sequence set forth in SEQ ID NO:8 relates to the polypeptide encoded by the wild-type *RSW1* gene sequences set forth in SEQ ID Nos:4 and 6.

The nucleotide sequence shown in SEQ ID NO:9 relates to a cellulose gene as hereinbefore defined which comprises a further wild-type variant of the *Arabidopsis thaliana RSW1* gene set forth in SEQ ID Nos:3 and 5. The nucleotide sequence variant is designated *Ath*-B. The

amino acid sequence set forth in SEQ ID NO:10 relates to the polypeptide encoded by the wild-type *RSW1* gene sequence set forth in SEQ ID No:9.

The nucleotide sequence shown in SEQ ID NO:11 relates to a cellulose gene as hereinbefore defined which comprises a cDNA equivalent of the *Arabidopsis thaliana rsw1* gene. The *rsw1* gene is a mutant cellulose gene which produces a radial root swelling phenotype as described by Baskin *et al* (1992). The present inventors have shown herein that the *rsw1* gene also produces reduced inflorescence length, reduced fertility, misshapen epidermal cells, reduced cellulose content and the accumulation of non-crystalline β-1,4-glucan, amongst others, when expressed in plant cells. The *rsw1* nucleotide sequence is a further variant of the nucleotide sequences set forth in SEQ ID Nos:3 and 5. The amino acid sequence set forth in SEQ ID NO:12 relates to the rsw1 polypeptide encoded by the mutant *rsw1* gene sequence set forth in SEQ ID No:11.

The nucleotide sequence shown in SEQ ID NO:13 relates to a cellulose gene as hereinbefore defined which comprises a cDNA equivalent of the *Oryza sativa RSW1* or *RSW1*-like gene. The nucleotide sequence is closely-related to the *Arabidopsis thaliana RSW1* and *rsw1* nucleotide sequences set forth herein (SEQ ID Nos:1, 3, 4, 5, 7, 9 and 11). The amino acid sequence set forth in SEQ ID NO:14 relates to the polypeptide encoded by the *RSW1* or *RSW1*-like gene sequences set forth in SEQ ID No:13.

Those skilled in the art will be aware of procedures for the isolation of further cellulose genes to those specifically described herein, for example further cDNA sequences and genomic gene equivalents, when provided with one or more of the nucleotide sequences set forth in SEQ ID Nos:1, 3, 4, 5, 7, 9, 11, or 13. In particular, hybridisations may be performed using one or more nucleic acid hybridisation probes comprising at least 10 contiguous nucleotides and preferably at least 50 contiguous nucleotides derived from the nucleotide sequences set forth herein, to isolate cDNA clones, mRNA molecules, genomic clones from a genomic library (in particular genomic clones containing the entire 5' upstream region of the gene including the promoter sequence, and the entire coding region and 3'-untranslated sequences), and/or

synthetic oligonucleotide molecules, amongst others. The present invention clearly extends to such related sequences.

The invention further extends to any homologues, analogues or derivatives of any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13.

A further aspect of the present invention contemplates a nucleic acid molecule which encodes or is complementary to a nucleic acid molecule which encodes, a polypeptide which is required for cellulose biosynthesis in a plant, such as cellulose synthase, and which is capable of hybridising under at least low stringency conditions to the nucleic acid molecule set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or to a complementary strand thereof.

As an exemplification of this embodiment, the present inventors have shown that it is possible to isolate variants of the *Arabidopsis thaliana RSW1* gene sequence set forth in SEQ ID NO:3, by hybridization under low stringency conditions. Such variants include related sequences derived from *Gossypium hirsutum* (cotton), *Eucalyptus ssp.* and *A. thaliana.* Additional variant are clearly encompassed by the present invention.

Preferably, the nucleic acid molecule further comprises a nucleotide sequence which encodes, or is complementary to a nucleotide sequence which encodes, a cellulose synthase polypeptide, more preferably a cellulose synthase which is associated with the primary or secondary plant cell wall of the plant species from which said nucleic acid molecule was derived.

More preferably, the nucleic acid molecule according to this aspect of the invention encodes or is complementary to a nucleic acid molecule which encodes, a polypeptide which is required for cellulose biosynthesis in a plant, such as cellulose synthase, and which is capable of hybridising under at least medium stringency conditions to the nucleic acid molecule set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or to a complementary

strand thereof.

Even more preferably, the nucleic acid molecule according to this aspect of the invention encodes or is complementary to a nucleic acid molecule which encodes, a polypeptide which is required for cellulose biosynthesis in a plant, such as cellulose synthase, and which is capable of hybridising under at least high stringency conditions to the nucleic acid molecule set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or to a complementary strand thereof.

For the purposes of defining the level of stringency, a low stringency is defined herein as being a hybridisation and/or a wash carried out in 6xSSC buffer, 0.1% (w/v) SDS at 28°C. Generally, the stringency is increased by reducing the concentration of SSC buffer, and/or increasing the concentration of SDS and/or increasing the temperature of the hybridisation and/or wash. A medium stringency comprises a hybridisation and/or a wash carried out in 0.2xSSC-2xSSC buffer, 0.1% (w/v) SDS at 42°C to 65°C, while a high stringency comprises a hybridisation and/or a wash carried out in 0.1xSSC-0.2xSSC buffer, 0.1% (w/v) SDS at a temperature of at least 55°C. Conditions for hybridisations and washes are well understood by one normally skilled in the art. For the purposes of further clarification only, reference to the parameters affecting hybridisation between nucleic acid molecules is found in pages 2.10.8 to 2.10.16. of Ausubel *et al.* (1987), which is herein incorporated by reference.

In an even more preferred embodiment of the invention, the isolated nucleic acid molecule further comprises a sequence of nucleotides which is at least 40% identical to at least 10 contiguous nucleotides derived from any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary strand thereof.

Still more preferably, the isolated nucleic acid molecule further comprises a sequence of nucleotides which is at least 40% identical to at least 50 contiguous nucleotides derived from the sequence set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary strand thereof.

The present invention is particularly directed to a nucleic acid molecule which is capable of functioning as a cellulose gene as hereinbefore defined, for example a cellulose synthase gene such as, but not limited to, the *Arabidopsis thaliana, Oryza sativa,* wheat, barley, maize, *Brassica ssp., Gossypium hirsutum* or *Eucalyptus ssp.* cellulose synthase genes, amongst others. The subject invention clearly contemplates additional cellulose genes to those specifically described herein which are derived from these plant species.

The invention further contemplates other sources of cellulose genes such as but not limited to, tissues and cultured cells of plant origin. Preferred plant species according to this embodiment include hemp, jute, flax and woody plants including, but not limited to *Pinus ssp., Populus ssp., Picea spp.*, amongst others.

A genetic sequence which encodes or is complementary to a sequence which encodes a polypeptide which is involved in cellulose biosynthesis may correspond to the naturally occurring sequence or may differ by one or more nucleotide substitutions, deletions and/or additions. Accordingly, the present invention extends to cellulose genes and any functional genes, mutants, derivatives, parts, fragments, homologues or analogues thereof or non-functional molecules but which are at least useful as, for example, genetic probes, or primer sequences in the enzymatic or chemical synthesis of said gene, or in the generation of immunologically interactive recombinant molecules.

In a particularly preferred embodiment, the cellulose genetic sequences are employed to identify and isolate similar genes from plant cells, tissues, or organ types of the same species, or from the cells, tissues, or organs of another plant species.

According to this embodiment, there is contemplated a method for identifying a related cellulose gene or related cellulose genetic sequence, for example a cellulose synthase or cellulose synthase-like gene, said method comprising contacting genomic DNA, or mRNA, or cDNA with a hybridisation effective amount of a first cellulose genetic sequence comprising any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary

sequence, homologue, analogue or derivative thereof derived from at least 10 contiguous nucleotides of said first sequence, and then detecting said hybridisation.

Preferably, the first genetic sequence comprises at least 50 contiguous nucleotides, even more preferably at least 100 contiguous nucleotides and even more preferably at least 500 contiguous nucleotides, derived from any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary strand, homologue, analogue or derivative thereof.

The related cellulose gene or related cellulose genetic sequence may be in a recombinant form, in a virus particle, bacteriophage particle, yeast cell, animal cell, or a plant cell. Preferably, the related cellulose gene or related cellulose genetic sequence is derived from a plant species, such as a monocotyledonous plant or a dicotyledonous plant selected from the list comprising *Arabidopsis thaliana*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, hemp, jute, flax, and woody plants including, but not limited to *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

More preferably, related cellulose gene or related cellulose genetic sequence is derived from a plant which is useful in the fibre or timber industries, for example *Gossypium hirsutum* (cotton), hemp, jute, flax, *Eucalyptus ssp.* or *Pinus ssp.*, amongst others. Alternatively, the related cellulose gene or related cellulose genetic sequence is derived from a plant which is useful in the cereal or starch industry, for example wheat, barley, rice or maize, amongst others.

In a particularly preferred embodiment, the first cellulose genetic sequence is labelled with a reporter molecule capable of giving an identifiable signal (e.g. a radioisotope such as ^{32}P or ^{35}S or a biotinylated molecule).

An alternative method contemplated in the present invention involves hybridising two nucleic acid "primer molecules" to a nucleic acid "template molecule" which comprises a related cellulose gene or related cellulose genetic sequence or a functional part thereof, wherein the

first of said primers comprises contiguous nucleotides derived from any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13 or a homologue, analogue or derivative thereof and the second of said primers comprises contiguous nucleotides complementary to any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13. Specific nucleic acid molecule copies of the template molecule are amplified enzymatically in a polymerase chain reaction, a technique that is well known to one skilled in the art.

In a preferred embodiment, each nucleic acid primer molecule is at least 10 nucleotides in length, more preferably at least 20 nucleotides in length, even more preferably at least 30 nucleotides in length, still more preferably at least 40 nucleotides in length and even still more preferably at least 50 nucleotides in length.

Furthermore, the nucleic acid primer molecules consists of a combination of any of the nucleotides adenine, cytidine, guanine, thymidine, or inosine, or functional analogues or derivatives thereof which are at least capable of being incorporated into a polynucleotide molecule without having an inhibitory effect on the hybridisation of said primer to the template molecule in the environment in which it is used.

Furthermore, one or both of the nucleic acid primer molecules may be contained in an aqueous mixture of other nucleic acid primer molecules, for example a mixture of degenerate primer sequences which vary from each other by one or more nucleotide substitutions or deletions. Alternatively, one or both of the nucleic acid primer molecules may be in a substantially pure form.

The nucleic acid template molecule may be in a recombinant form, in a virus particle, bacteriophage particle, yeast cell, animal cell, or a plant cell. Preferably, the nucleic acid

template molecule is derived from a plant cell, tissue or organ, in particular a cell, tissue or organ derived from a plant selected from the list comprising *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, and woody plants including, but not limited to *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

Those skilled in the art will be aware that there are many known variations of the basic polymerase chain reaction procedure, which may be employed to isolate a related cellulose gene or related cellulose genetic sequence when provided with the nucleotide sequences set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13. Such variations are discussed, for example, in McPherson *et al* (1991). The present invention extends to the use of all such variations in the isolation of related cellulose genes or related cellulose genetic sequences using the nucleotide sequences embodied by the present invention.

The isolated nucleic acid molecule according to any of the further embodiments may be cloned into a plasmid or bacteriophage molecule, for example to facilitate the preparation of primer molecules or hybridisation probes or for the production of recombinant gene products. Methods for the production of such recombinant plasmids, cosmids, bacteriophage molecules or other recombinant molecules are well-known to those of ordinary skill in the art and can be accomplished without undue experimentation. Accordingly, the invention further extends to any recombinant plasmid, bacteriophage, cosmid or other recombinant molecule comprising the nucleotide sequence set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary sequence, homologue, analogue or derivative thereof.

The nucleic acid molecule of the present invention is also useful for developing genetic constructs which express a cellulose genetic sequence, thereby providing for the increased expression of genes involved in cellulose biosynthesis in plants, selected for example from the list comprising *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, and woody plants including, but not limited to *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others. The present invention

particularly contemplates the modification of cellulose biosynthesis in cotton, hemp, jute, flax, *Eucalyptus ssp.* and *Pinus ssp.*, amongst others.

The present inventors have discovered that the genetic sequences disclosed herein are capable of being used to modify the level of non-crystalline β-1,4,-glucan, in addition to altering cellulose levels when expressed, particularly when expressed in plants cells. In particular, the *Arabidopsis thaliana rsw*1 mutant has increased levels of non-crystalline β-1,4,-glucan, when grown at 31°C, compared to wild-type plants, grown under identical conditions. The expression of a genetic sequence described herein in the antisense orientation in transgenic plants grown at only 21°C is shown to reproduce many aspects of the *rsw*1 mutant phenotype.

Accordingly, the present invention clearly extends to the modification of non-crystalline β-1,4,-glucan biosynthesis in plants, selected for example from the list comprising *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, and woody plants including, but not limited to *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others. The present invention particularly contemplates the modification of non-crystalline β-1,4,-glucan biosynthesis in cotton, hemp, jute, flax, *Eucalyptus ssp.* and *Pinus ssp.*, amongst others.

The present invention further extends to the production and use of non-crystalline β-1,4-glucan and to the use of the glucan to modify the properties of plant cell walls or cotton fibres or wood fibres. Such modified properties are described herein (Example 13).

The inventors have discovered that the *rsw*1 mutant has altered carbon partitioning compared to wild-type plants, resulting in significantly higher starch levels therein. The isolated nucleic acid molecules provided herein are further useful for altering the carbon partitioning in a cell. In particular, the present invention contemplates increased starch production in transgenic plants expressing the nucleic acid molecule of the invention in the antisense orientation or alterntively, expressing a ribozyme or co-suppression molecule comprising the nucleic acid sequence of the invention.

The invention further contemplates reduced starch and/or non-crystalline β-1,4-glucan product in transgenic plants expressing the nucleic acid molecule of the invention in the sense orientation such that cellulose production is increased therein.

Wherein it is desired to increase cellulose production in a plant cell, the coding region of a cellulose gene is placed operably behind a promoter, in the sense orientation, such that a cellulose gene product is capable of being expressed under the control of said promoter sequence. In a preferred embodiment, the cellulose genetic sequence is a cellulose synthase genomic sequence, cDNA molecule or protein-coding sequence.

In a particularly preferred embodiment, the cellulose genetic sequence comprises a sequence of nucleotides substantially the same as the sequence set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13 or a homologue, analogue or derivative thereof.

Wherein it is desirable to reduce the content of cellulose or to increase the content of non-crystalline β-1,4-glucan, the nucleic acid molecule of the present invention is expressed in the antisense orientation under the control of a suitable promoter. Additionally, the nucleic acid molecule of the invention is also useful for developing ribozyme molecules, or in co-suppression of a cellulose gene. The expression of an antisense, ribozyme or co-suppression molecule comprising a cellulose gene, in a cell such as a plant cell, fungal cell, insect cell. animal cell, yeast cell or bacterial cell, may also increase the solubility, digestibility or extractability of metabolites from plant tissues or alternatively, or increase the availability of carbon as a precursor for any secondary metabolite other than cellulose (e.g. starch or sucrose). By targeting the endogenous cellulose gene, expression is diminished, reduced or otherwise lowered to a level that results in reduced deposition of cellulose in the primary or secondary cell walls of the plant cell, fungal cell, insect cell. animal cell, yeast cell or bacterial cell, and more particularly, a plant cell. Additionally, or alternatively, the content of non-crystalline β-1,4-glucan is increased in such cells.

Co-suppression is the reduction in expression of an endogenous gene that occurs when one

or more copies of said gene, or one or more copies of a substantially similar gene are introduced into the cell. The present invention also extends to the use of co-suppression to inhibit the expression of a gene which encodes a cellulose gene product, such as but not limited to cellulose synthase. Preferably, the co-suppression molecule of the present invention targets a plant mRNA molecule which encodes a cellulose synthase enzyme, for example a plant, fungus, or bacterial cellulose synthase mRNA, and more preferably a plant mRNA derived from *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, or a woody plant such as *Pinus ssp.*, *Populus ssp.*, or *Picea spp.*, amongst others.

In a particularly preferred embodiment, the gene which is targeted by a co-suppression molecule, comprises a sequence of nucleotides set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complement, homologue, analogue or derivative thereof.

In the context of the present invention, an antisense molecule is an RNA molecule which is transcribed from the complementary strand of a nuclear gene to that which is normally transcribed to produce a "sense" mRNA molecule capable of being translated into a polypeptide component of the cellulose biosynthetic pathway. The antisense molecule is therefore complementary to the mRNA transcribed from a sense cellulose gene or a part thereof. Although not limiting the mode of action of the antisense molecules of the present invention to any specific mechanism, the antisense RNA molecule possesses the capacity to form a double-stranded mRNA by base pairing with the sense mRNA, which may prevent translation of the sense mRNA and subsequent synthesis of a polypeptide gene product.

Preferably, the antisense molecule of the present invention targets a plant mRNA molecule which encodes a cellulose gene product, for example cellulose synthase. Preferably, the antisense molecule of the present invention targets a plant mRNA molecule which encodes a cellulose synthase enzyme, for example a plant mRNA derived from *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, or a woody plant such as *Pinus ssp.*, *Populus ssp.*, or *Picea spp.*, amongst

others.

In a particularly preferred embodiment, the antisense molecule of the invention targets an mRNA molecule encoded by any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a homologue, analogue or derivative thereof.

Ribozymes are synthetic RNA molecules which comprise a hybridising region complementary to two regions, each of at least 5 contiguous nucleotide bases in the target sense mRNA. In addition, ribozymes possess highly specific endoribonuclease activity, which autocatalytically cleaves the target sense mRNA. A complete description of the function of ribozymes is presented by Haseloff and Gerlach (1988) and contained in International Patent Application No. WO89/05852.

The present invention extends to ribozyme which target a sense mRNA encoding a cellulose gene product, thereby hybridising to said sense mRNA and cleaving it, such that it is no longer capable of being translated to synthesise a functional polypeptide product. Preferably, the ribozyme molecule of the present invention targets a plant mRNA molecule which encodes a cellulose synthase enzyme, for example a plant mRNA derived from *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, hemp, jute, flax, or a woody plant such as *Pinus ssp.*, *Populus ssp.*, or *Picea spp.*, amongst others.

In a particularly preferred embodiment, the ribozyme molecule will target an mRNA encoded by any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a homologue, analogue or derivative thereof.

According to this embodiment, the present invention provides a ribozyme or antisense molecule comprising at least 5 contiguous nucleotide bases derived from any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary nucleotide sequence or a homologue, analogue or derivative thereof, wherein said antisense or ribozyme molecule is able to form a hydrogen-bonded complex with a sense mRNA encoding a cellulose gene

product to reduce translation thereof.

In a preferred embodiment, the antisense or ribozyme molecule comprises at least 10 to 20 contiguous nucleotides derived from any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a complementary nucleotide sequence or a homologue, analogue or derivative thereof. Although the preferred antisense and/or ribozyme molecules hybridise to at least about 10 to 20 nucleotides of the target molecule, the present invention extends to molecules capable of hybridising to at least about 50-100 nucleotide bases in length, or a molecule capable of hybridising to a full-length or substantially full-length mRNA encoded by a cellulose gene, such as a cellulose synthase gene.

Those skilled in the art will be aware of the necessary conditions, if any, for selecting or preparing the antisense or ribozyme molecules of the invention.

It is understood in the art that certain modifications, including nucleotide substitutions amongst others, may be made to the antisense and/or ribozyme molecules of the present invention, without destroying the efficacy of said molecules in inhibiting the expression of a gene encoding a cellulose gene product such as cellulose synthase. It is therefore within the scope of the present invention to include any nucleotide sequence variants, homologues, analogues, or fragments of the said gene encoding same, the only requirement being that said nucleotide sequence variant, when transcribed, produces an antisense and/or ribozyme molecule which is capable of hybridising to a sense mRNA molecule which encodes a cellulose gene product.

Gene targeting is the replacement of an endogenous gene sequence within a cell by a related DNA sequence to which it hybridises, thereby altering the form and/or function of the endogenous gene and the subsequent phenotype of the cell. According to this embodiment, at least a part of the DNA sequence defined by any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a related cellulose genetic sequence, may be introduced into target cells containing an endogenous cellulose gene, thereby replacing said endogenous cellulose gene.

According to this embodiment, the polypeptide product of said cellulose genetic sequence possesses different catalytic activity and/or expression characteristics, producing in turn modified cellulose deposition in the target cell. In a particularly preferred embodiment of the invention, the endogenous cellulose gene of a plant is replaced with a gene which is merely capable of producing non-crystalline β-1,4-glucan polymers or alternatively which is capable of producing a modified cellulose having properties similar to synthetic fibres such as rayon, in which the β-1,4-glucan polymers are arranged in an antiparallel configuration relative to one another.

The present invention extends to genetic constructs designed to facilitate expression of a cellulose genetic sequence which is identical, or complementary to the sequence set forth in any one or more of SEQ ID Nos:1, 3, 4, 5, 7, 9, 11 or 13, or a functional derivative, part, homologue, or analogue thereof, or a genetic construct designed to facilitate expression of a sense molecule, an antisense molecule, ribozyme molecule, co-suppression molecule, or gene targeting molecule containing said genetic sequence.

The said genetic construct of the present invention comprises the foregoing sense, antisense, or ribozyme, or co-suppression nucleic acid molecule, or gene-targeting molecule, placed operably under the control of a promoter sequence capable of regulating the expression of the said nucleic acid molecule in a prokaryotic or eukaryotic cell, preferably a plant cell. The said genetic construct optionally comprises, in addition to a promoter and sense, or antisense, or ribozyme, or co-suppression, or gene-targeting nucleic acid molecule, a terminator sequence.

The term "terminator" refers to a DNA sequence at the end of a transcriptional unit which signals termination of transcription. Terminators are 3'-non-translated DNA sequences containing a polyadenylation signal, which facilitates the addition of polyadenylate sequences to the 3'-end of a primary transcript. Terminators active in plant cells are known and described in the literature. They may be isolated from bacteria, fungi, viruses, animals and/or plants. Examples of terminators particularly suitable for use in the genetic constructs

of the present invention include the nopaline synthase (NOS) gene terminator of *Agrobacterium tumefaciens*, the terminator of the Cauliflower mosaic virus (CaMV) 35S gene, and the *zein* gene terminator from *Zea mays*.

Reference herein to a "promoter" is to be taken in its broadest context and includes the transcriptional regulatory sequences of a classical genomic gene, including the TATA box which is required for accurate transcription initiation, with or without a CCAAT box sequence and additional regulatory elements (i.e. upstream activating sequences, enhancers and silencers) which alter gene expression in response to developmental and/or external stimuli, or in a tissue-specific manner. A promoter is usually, but not necessarily, positioned upstream or 5', of a structural gene, the expression of which it regulates. Furthermore, the regulatory elements comprising a promoter are usually positioned within 2 kb of the start site of transcription of the gene.

In the present context, the term "promoter" is also used to describe a synthetic or fusion molecule, or derivative which confers, activates or enhances expression of said sense, antisense, or ribozyme, or co-suppression nucleic acid molecule, in a plant cell. Preferred promoters may contain additional copies of one or more specific regulatory elements, to further enhance expression of a sense antisense, ribozyme or co-suppression molecule and/or to alter the spatial expression and/or temporal expression of said sense or antisense, or ribozyme, or co-suppression, or gene-targeting molecule. For example, regulatory elements which confer copper inducibility may be placed adjacent to a heterologous promoter sequence driving expression of a sense, or antisense, or ribozyme, or co-suppression, or gene-targeting molecule, thereby conferring copper inducibility on the expression of said molecule.

Placing a sense or ribozyme, or antisense, or co-suppression, or gene-targeting molecule under the regulatory control of a promoter sequence means positioning the said molecule such that expression is controlled by the promoter sequence. Promoters are generally positioned 5' (upstream) to the genes that they control. In the construction of heterologous promoter/structural gene combinations it is generally preferred to position the promoter at a

distance from the gene transcription start site that is approximately the same as the distance between that promoter and the gene it controls in its natural setting, i.e., the gene from which the promoter is derived. As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting, i.e., the genes from which it is derived. Again, as is known in the art, some variation in this distance can also occur.

Examples of promoters suitable for use in genetic constructs of the present invention include viral, fungal, bacterial, animal and plant derived promoters capable of functioning in prokaryotic or eukaryotic cells. Preferred promoters are those capable of regulating the expression of the subject cellulose genes of the innvention in plants cells, fungal cells, insect cells, yeast cells, animal cells or bacterial cells, amongst others. Particularly preferred promoters are capable of regulating expression of the subject nucleic acid molecules in plant cells. The promoter may regulate the expression of the said molecule constitutively, or differentially with respect to the tissue in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, or plant pathogens, or metal ions, amongst others. Preferably, the promoter is capable of regulating expression of a sense, or ribozyme, or antisense, or co-suppression molecule or gene targeting, in a plant cell. Examples of preferred promoters include the CaMV 35S promoter, NOS promoter, octopine synthase (OCS) promoter and the like.

In a most preferred embodiment, the promoter is capable of expression in any plant cell, such as, but not limited to a plant selected from the list comprising *Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, and woody plants including, but not limited to *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

In a particularly preferred embodiment, the promoter may be derived from a genomic clone

encoding a cellulose gene product, in particular the promoter contained in the sequence set forth in SEQ ID NO:3 or SEQ ID NO:4. Preferably, the promoter sequence comprises nucleotide 1 to about 1900 of SEQ ID NO:3 or nucleotides 1 to about 700 of SEQ ID NO:4 or a homologue, analogue or derivative capable of hybridizing thereto under at least low stringency conditions.

Optionally, the genetic construct of the present invention further comprises a terminator sequence.

In an exemplification of this embodiment, there is provided a binary genetic construct comprising the isolated nucleotide sequence of nucleotides set forth in SEQ ID NO:3. There is also provided a genetic construct comprising the isolated nucleotide sequence of nucleotides set forth in SEQ ID NO:1, in the antisense orientation, placed operably in connection with the CaMV 35S promoter.

In the present context, the term "in operable connection with" means that expression of the isolated nucleotide sequence is under the control of the promoter sequence with which it is connected, regardless of the relative physical distance of the sequences from each other or their relative orientation with respect to each other.

An alternative embodiment of the invention is directed to a genetic construct comprising a promoter or functional derivative, part, fragment, homologue, or analogue thereof, which is capable of directing the expression of a polypeptide early in the development of a plant cell at a stage when the cell wall is developing, such as during cell expansion or during cell division. In a particularly preferred embodiment, the promoter is contained in the sequence set forth in SEQ ID NO:3 or SEQ ID NO:4. Preferably, the promoter sequence comprises nucleotide 1 to about 1900 of SEQ ID NO:3 or nucleotides 1 to about 700 of SEQ ID NO:4 or a homologue, analogue or derivative capable of hybridizing thereto under at least low stringency conditions.

The polypeptide may be a reporter molecule which is encoded by a gene such as the bacterial β-glucuronidase gene or chloramphenicol acetyltransferase gene or alternatively, the firefly luciferase gene. Alternatively, the polypeptide may be encoded by a gene which is capable of producing a modified cellulose in the plant cell when placed in combination with the normal complement of cellulose genes which are expressible therein, for example it may be a cellulose-like gene obtained from a bacterial or fungal source or a cellulose gene obtained from a plant source.

The genetic constructs of the present invention are particularly useful in the production of crop plants with altered cellulose content or structure. In particular, the rate of cellulose deposition may be reduced leading to a reduction in the total cellulose content of plants by transferring one or more of the antisense, ribozyme or co-suppression molecules described *supra* into a plant or alternatively, the same or similar end-result may be achieved by replacing an endogenous cellulose gene with an inactive or modified cellulose gene using gene-targeting approaches. The benefits to be derived from reducing cellulose content in plants are especially apparent in food and fodder crops such as, but not limited to maize, wheat, barley, rye, rice, barley, millet or sorghum, amongst others where improved digestibility of said crop is desired. The foregoing antisense, ribozyme or co-suppression molecules are also useful in producing plants with altered carbon partitioning such that increased carbon is available for growth, rather than deposited in the form of cellulose.

Alternatively, the introduction to plants of additional copies of a cellulose gene in the 'sense' orientation and under the control of a strong promoter is useful for the production of plants with increased cellulose content or more rapid rates of cellulose biosynthesis. Accordingly, such plants may exhibit a range of desired traits including, but not limited to modified strength and/or shape and/or properties of fibres, cell and plants, increased protection against chemical, physical or environmental stresses such as dehydration, heavy metals (e.g. cadmium) cold, heat or wind, increased resistance to attack by pathogens such as insects, nematodes and the like which physically penetrate the cell wall barrier during invasion/infection of the plant.

Alternatively, the production of plants with altered physical properties is made possible by the introduction thereto of altered cellulose gene(s). Such plants may produce β-1,4-glucan which is either non-crystalline or shows altered crystallinity. Such plants may also exhibit a range of desired traits including but not limited to, altered dietary fibre content, altered digestibility and degradability or producing plants with altered extractability properties.

Furthermore, genetic constructs comprising a plant cellulose gene in the 'sense' orientation may be used to complement the existing range of cellulose genes present in a plant, thereby altering the composition or timing of deposition of cellulose deposited in the cell wall of said plant. In a preferred embodiment, the cellulose gene from one plant species or a β-1,4-glucan synthase gene from a non-plant species is used to transform a plant of a different species, thereby introducing novel cellulose biosynthetic metabolism to the second-mentioned plant species.

In a related embodiment, a recombinant fusion polypeptide may be produced containing the active site from one cellulose gene product fused to another cellulose gene product, wherein said fusion polypeptide exhibits novel catalytic properties compared to either 'parent' polypeptide from which it is derived. Such fusion polypeptides may be produced by conventional recombinant DNA techniques known to those skilled in the art, either by introducing a recombinant DNA capable of expressing the entire fusion polypeptide into said plant or alternatively, by a gene-targeting approach in which recombination at the DNA level occurs *in vivo* and the resultant gene is capable of expressing a recombinant fusion polypeptide.

The present invention extends to all transgenic methods and products described *supra*, including genetic constructs.

The recombinant DNA molecule carrying the sense, antisense, ribozyme or co-suppression molecule of the present invention and/or genetic construct comprising the same, may be introduced into plant tissue, thereby producing a "transgenic plant", by various techniques

known to those skilled in the art. The technique used for a given plant species or specific type of plant tissue depends on the known successful techniques. Means for introducing recombinant DNA into plant tissue include, but are not limited to, transformation (Paszkowski *et al.*, 1984), electroporation (Fromm *et al.*, 1985), or microinjection of the DNA (Crossway *et al.*, 1986), or T-DNA-mediated transfer from *Agrobacterium* to the plant tissue. Representative T-DNA vector systems are described in the following references: An *et al.*(1985); Herrera-Estrella *et al.* (1983a,b); Herrera-Estrella *et al.* (1985). Once introduced into the plant tissue, the expression of the introduced gene may be assayed in a transient expression system, or it may be determined after selection for stable integration within the plant genome. Techniques are known for the *in vitro* culture of plant tissue, and in a number of cases, for regeneration into whole plants. Procedures for transferring the introduced gene from the originally transformed plant into commercially useful cultivars are known to those skilled in the art.

A still further aspect of the present invention extends to a transgenic plant such as a crop plant, carrying the foregoing sense, antisense, ribozyme, co-suppression, or gene-targeting molecule and/or genetic constructs comprising the same. Preferably, the transgenic plant is one or more of the following:*Arabidopsis thaliana*, *Oryza sativa*, wheat, barley, maize, *Brassica ssp.*, *Gossypium hirsutum* and *Eucalyptus ssp.*, hemp, jute, flax, *Pinus ssp.*, *Populus ssp.*, or *Picea spp.* Additional species are not excluded.

The present invention further extends to the progeny of said transgenic plant.

Yet another aspect of the present invention provides for the expression of the subject genetic sequence in a suitable host (e.g. a prokaryote or eukaryote) to produce full length or non-full length recombinant cellulose gene products.

Hereinafter the term "cellulose gene product" shall be taken to refer to a recombinant product of a cellulose gene as hereinbefore defined. Accordingly, the term "cellulose gene product" includes a polypeptide product of any gene involved in the cellulose biosynthetic pathway in

plants, such as, but not limited to a cellulose synthase gene product.

Preferably, the recombinant cellulose gene product comprises an amino acid sequence having the catalytic activity of a cellulose synthase polypeptide or a functional mutant, derivative part, fragment, or analogue thereof.

In a particularly preferred embodiment of the invention, the recombinant cellulose gene product comprises a sequence or amino acids that is at least 40% identical to any one or more of SEQ ID Nos:2, 6, 8, 10, 12 or 14, or a homologue, analogue or derivative thereof.

Single and three-letter abbreviations used for amino acid residues contained in the specification are provided in Table 1.

In the present context, "homologues" of an amino acid sequence refer to those polypeptides, enzymes or proteins which have a similar catalytic activity to the amino acid sequences described herein, notwithstanding any amino acid substitutions, additions or deletions thereto. A homologue may be isolated or derived from the same or another plant species as the species from which the polypeptides of the invention are derived.

"Analogues" encompass polypeptides of the invention notwithstanding the occurrence of any non-naturally occurring amino acid analogues therein.

"Derivatives" include modified peptides in which ligands are attached to one or more of the amino acid residues contained therein, such as carbohydrates, enzymes, proteins, polypeptides or reporter molecules such as radionuclides or fluorescent compounds. Glycosylated, fluorescent, acylated or alkylated forms of the subject peptides are particularly contemplated by the present invention. Additionally, derivatives of an amino acid sequence described herein which comprises fragments or parts of the subject amino acid sequences are within the scope of the invention, as are homopolymers or heteropolymers comprising two or more copies of the subject polypeptides. Procedures for derivatizing peptides are well-known in the

art.

TABLE 1

Amino Acid	Three-letter Abbreviation	One-letter Symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
D-alanine	Dal	X
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tryosine	Tyr	Y
Valine	Val	V
Any amino acid	Xaa	X

Substitutions encompass amino acid alterations in which an amino acid is replaced with a different naturally-occurring or a non-conventional amino acid residue. Such substitutions may be classified as "conservative", in which an amino acid residue contained in a cellulose gene product is replaced with another naturally-occurring amino acid of similar character, for example Gly↔Ala, Val↔Ile↔Leu, Asp↔Glu, Lys↔Arg, Asn↔Gln or Phe↔Trp↔Tyr.

Substitutions encompassed by the present invention may also be "non-conservative", in which an amino acid residue which is present in a cellulose gene product described herein is substituted with an amino acid with different properties, such as a naturally-occurring amino acid from a different group (eg. substituted a charged or hydrophobic amino acid with alanine), or alternatively, in which a naturally-occurring amino acid is substituted with a non-conventional amino acid.

Non-conventional amino acids encompassed by the invention include, but are not limited to those listed in Table 2.

Amino acid substitutions are typically of single residues, but may be of multiple residues, either clustered or dispersed.

Amino acid deletions will usually be of the order of about 1-10 amino acid residues, while insertions may be of any length. Deletions and insertions may be made to the N-terminus, the C-terminus or be internal deletions or insertions. Generally, insertions within the amino acid sequence will be smaller than amino- or carboxy-terminal fusions and of the order of 1-4 amino acid residues.

A homologue, analogue or derivative of a cellulose gene product as referred to herein may readily be made using peptide synthetic techniques well-known in the art, such as solid phase peptide synthesis and the like, or by recombinant DNA manipulations. Techniques for making substituent mutations at pre-determined sites using recombinant DNA technology, for example by M13 mutagenesis, are also well-known. The manipulation of nucleic acid

molecules to produce variant peptides, polypeptides or proteins which manifest as substitutions, insertions or deletions are well-known in the art.

The cellulose gene products described herein may be derivatized further by the inclusion or attachment thereto of a protective group which prevents, inhibits or slows proteolytic or cellular degradative processes. Such derivatization may be useful where the half-life of the subject polypeptide is required to be extended, for ample to increase the amount of cellulose produced in a primary or secondary cell wall of a plant cell or alternatively, to increase the amount of protein produced in a bacterial or eukaryotic expression system. Examples of chemical groups suitable for this purpose include, but are not limited to, any of the non-conventional amino acid residues listed in Table 2, in particular a D-stereoisomer or a methylated form of a naturally-occurring amino acid listed in Table 1. Additional chemical groups which are useful for this purpose are selected from the list comprising aryl or heterocyclic N-acyl substituents, polyalkylene oxide moieties, desulphatohirudin muteins, alpha-muteins, alpha-aminophosphonic acids, water-soluble polymer groups such as polyethylene glycol attached to sugar residues using hydrazone or oxime groups, benzodiazepine dione derivatives, glycosyl groups such as beta-glycosylamine or a derivative thereof, isocyanate conjugated to a polyol functional group or polyoxyethylene polyol capped with diisocyanate, amongst others. Similarly, a cellulose gene product or a homologue, analogue or derivative thereof may be cross-linked or fused to itself or to a protease inhibitor peptide, to reduce susceptibility of said molecule to proteolysis.

TABLE 2

Non-conventional amino acid	Code	Non-conventional amino acid	Code
α-aminobutyric acid	Abu	L-N-methylalanine	Nmala
α-amino-α-methylbutyrate	Mgabu	L-N-methylarginine	Nmarg
aminocyclopropane-carboxylate	Cpro	L-N-methylasparagine	Nmasn
		L-N-methylaspartic acid	Nmasp
aminoisobutyric acid	Aib	L-N-methylcysteine	Nmcys
aminonorbornyl-carboxylate	Norb	L-N-methylglutamine	Nmgln
		L-N-methylglutamic acid	Nmglu
cyclohexylalanine	Chexa	L-N-methylhistidine	Nmhis
cyclopentylalanine	Cpen	L-N-methylisolleucine	Nmile
D-alanine	Dal	L-N-methylleucine	Nmleu
D-arginine	Darg	L-N-methyllysine	Nmlys
D-aspartic acid	Dasp	L-N-methylmethionine	Nmmet
D-cysteine	Dcys	L-N-methylnorleucine	Nmnle
D-glutamine	Dgln	L-N-methylnorvaline	Nmnva
D-glutamic acid	Dglu	L-N-methylornithine	Nmorn
D-histidine	Dhis	L-N-methylphenylalanine	Nmphe
D-isoleucine	Dile	L-N-methylproline	Nmpro
D-leucine	Dleu	L-N-methylserine	Nmser
D-lysine	Dlys	L-N-methylthreonine	Nmthr
D-methionine	Dmet	L-N-methyltryptophan	Nmtrp
D-ornithine	Dorn	L-N-methyltyrosine	Nmtyr
D-phenylalanine	Dphe	L-N-methylvaline	Nmval
D-proline	Dpro	L-N-methylethylglycine	Nmetg
D-serine	Dser	L-N-methyl-t-butylglycine	Nmtbug
D-threonine	Dthr	L-norleucine	Nle

	D-tryptophan	Dtrp	L-norvaline	Nva
	D-tyrosine	Dtyr	α-methyl-aminoisobutyrate	Maib
	D-valine	Dval	α-methyl-γ-aminobutyrate	Mgabu
	D-α-methylalanine	Dmala	α-methylcyclohexylalanine	Mchexa
5	D-α-methylarginine	Dmarg	α-methylcylcopentylalanine	Mcpen
	D-α-methylasparagine	Dmasn	α-methyl-α-napthylalanine	Manap
	D-α-methylaspartate	Dmasp	α-methylpenicillamine	Mpen
	D-α-methylcysteine	Dmcys	N-(4-aminobutyl)glycine	Nglu
	D-α-methylglutamine	Dmgln	N-(2-aminoethyl)glycine	Naeg
10	D-α-methylhistidine	Dmhis	N-(3-aminopropyl)glycine	Norn
	D-α-methylisoleucine	Dmile	N-amino-α-methylbutyrate	Nmaabu
	D-α-methylleucine	Dmleu	α-napthylalanine	Anap
	D-α-methyllysine	Dmlys	N-benzylglycine	Nphe
	D-α-methylmethionine	Dmmet	N-(2-carbamylethyl)glycine	Ngln
15	D-α-methylornithine	Dmorn	N-(carbamylmethyl)glycine	Nasn
	D-α-methylphenylalanine	Dmphe	N-(2-carboxyethyl)glycine	Nglu
	D-α-methylproline	Dmpro	N-(carboxymethyl)glycine	Nasp
	D-α-methylserine	Dmser	N-cyclobutylglycine	Ncbut
	D-α-methylthreonine	Dmthr	N-cycloheptylglycine	Nchep
20	D-α-methyltryptophan	Dmtrp	N-cyclohexylglycine	Nchex
	D-α-methyltyrosine	Dmty	N-cyclodecylglycine	Ncdec
	D-α-methylvaline	Dmval	N-cylcododecylglycine	Ncdod
	D-N-methylalanine	Dnmala	N-cyclooctylglycine	Ncoct
	D-N-methylarginine	Dnmarg	N-cyclopropylglycine	Ncpro
25	D-N-methylasparagine	Dnmasn	N-cycloundecylglycine	Ncund
	D-N-methylaspartate	Dnmasp	N-(2,2-diphenylethyl)glycine	Nbhm
	D-N-methylcysteine	Dnmcys	N-(3,3-diphenylpropyl)glycine	Nbhe
	D-N-methylglutamine	Dnmgln	N-(3-guanidinopropyl)glycine	Narg
	D-N-methylglutamate	Dnmglu	N-(1-hydroxyethyl)glycine	Nthr
30	D-N-methylhistidine	Dnmhis	N-(hydroxyethyl))glycine	Nser

D-N-methylisoleucine	Dnmile	N-(imidazolylethyl))glycine	Nhis
D-N-methylleucine	Dnmleu	N-(3-indolylyethyl)glycine	Nhtrp
D-N-methyllysine	Dnmlys	N-methyl-γ-aminobutyrate	Nmgabu
N-methylcyclohexylalanine	Nmchexa	D-N-methylmethionine	Dnmmet
D-N-methylornithine	Dnmorn	N-methylcyclopentylalanine	Nmcpen
N-methylglycine	Nala	D-N-methylphenylalanine	Dnmphe
N-methylaminoisobutyrate	Nmaib	D-N-methylproline	Dnmpro
N-(1-methylpropyl)glycine	Nile	D-N-methylserine	Dnmser
N-(2-methylpropyl)glycine	Nleu	D-N-methylthreonine	Dnmthr
D-N-methyltryptophan	Dnmtrp	N-(1-methylethyl)glycine	Nval
D-N-methyltyrosine	Dnmtyr	N-methyla-napthylalanine	Nmanap
D-N-methylvaline	Dnmval	N-methylpenicillamine	Nmpen
γ-aminobutyric acid	Gabu	N-(*p*-hydroxyphenyl)glycine	Nhtyr
L-*t*-butylglycine	Tbug	N-(thiomethyl)glycine	Ncys
L-ethylglycine	Etg	penicillamine	Pen
L-homophenylalanine	Hphe	L-α-methylalanine	Mala
L-α-methylarginine	Marg	L-α-methylasparagine	Masn
L-α-methylaspartate	Masp	L-α-methyl-*t*-butylglycine	Mtbug
L-α-methylcysteine	Mcys	L-methylethylglycine	Metg
L-α-methylglutamine	Mgln	L-α-methylglutamate	Mglu
L-α-methylhistidine	Mhis	L-α-methylhomophenylalanine	Mhphe
L-α-methylisoleucine	Mile	N-(2-methylthioethyl)glycine	Nmet
L-α-methylleucine	Mleu	L-α-methyllysine	Mlys
L-α-methylmethionine	Mmet	L-α-methylnorleucine	Mnle
L-α-methylnorvaline	Mnva	L-α-methylornithine	Morn
L-α-methylphenylalanine	Mphe	L-α-methylproline	Mpro
L-α-methylserine	Mser	L-α-methylthreonine	Mthr
L-α-methyltryptophan	Mtrp	L-α-methyltyrosine	Mtyr
L-α-methylvaline	Mval	L-N-methylhomophenylalanine	Nmhphe

N-(N-(2,2-diphenylethyl) carbamylmethyl)glycine	Nnbhm	N-(N-(3,3-diphenylpropyl) carbamylmethyl)glycine	Nnbhe
1-carboxy-1-(2,2-diphenyl-ethylamino)cyclopropane	Nmbc		

In an alternative embodiment of the invention, the recombinant cellulose gene product is characterised by at least one functional β-glycosyl transferase domain contained therein.

The term "β-glycosyl transferase domain" as used herein refers to a sequence of amino acids which is highly conserved in different processive enzymes belonging to the class of glycosyl transferase enzymes (Saxena *et al.*, 1995), for example the bacterial β-1,4-glycosyl transferase enzymes and plant cellulose synthase enzymes amongst others, wherein said domain possesses a putative function in contributing to or maintaining the overall catalytic activity, substrate specificity or substrate binding of an enzyme in said enzyme class. The β-glycosyl transferase domain is recognisable by the occurrence of certain amino acid residues at particular locations in a polypeptide sequence, however there is no stretch of contiguous amino acid residues comprised therein.

As a consequence of the lack of contiguity in a β-glycosyl transferase domain, it is not a straightforward matter to isolate a cellulose gene by taking advantage of the presence of a β-glycosyl transferase domain in the polypeptide encoded by said gene. For example, the β-glycosyl transferase domain would not be easily utilisable as a probe to facilitate the rapid isolation of all β-glycosyl transferase genetic sequences from a particular organism and then to isolate from those genetic sequences a cellulose gene such as cellulose synthase.

In a preferred embodiment, the present invention provides an isolated polypeptide which:

(i)contains at least one structural β-glycosyl transferase domain as hereinbefore defined; and

(ii) has at least 40% amino acid sequence similarity to at least 20 contiguous amino acid residues set forth in any one or more of SEQ ID Nos:2, 6, 8, 10, 12 or 14, or a homologue, analogue or derivative thereof.

More preferably, the polypeptide of the invention is at least 40% identical to at least 50 contiguous amino acid residues, even more preferably at least 100 amino acid residues of any one or more of SEQ ID Nos:2, 6, 8, 10, 12 or 14, or a homologue, analogue or derivative thereof.

In a particularly preferred embodiment, the percentage similarity to any one or more of SEQ ID Nos:2, 6, 8, 10, 12 or 14 is at least 50-60%, more preferably at least 65-70%, even more preferably at least 75-80% and even more preferably at least 85-90%, including about 91% or 95%.

In a related embodiment, the present invention provides a "sequencably pure" form of the amino acid sequence described herein. "Sequencably pure" is hereinbefore described as substantially homogeneous to facilitate amino acid determination.

In a further related embodiment, the present invention provides a "substantially homogeneous" form of the subject amino acid sequence, wherein the term "substantially homogeneous" is hereinbefore defined as being in a form suitable for interaction with an immunologically interactive molecule. Preferably, the polypeptide is at least 20% homogeneous, more preferably at least 50% homogeneous, still more preferably at least 75% homogeneous and yet still more preferably at least about 95-100% homogenous, in terms of activity per microgram of total protein in the protein preparation.

The present invention further extends to a synthetic peptide of at least 5 amino acid residues in length derived from or comprising a part of the amino acid sequence set forth in any one or more of SEQ ID Nos:2, 6, 8, 10, 12 or 14, or having at least 40% similarity thereto.

Those skilled in the art will be aware that such synthetic peptides may be useful in the production of immunologically interactive molecules for the preparation of antibodies or as the peptide component of an immunoassay.

The invention further extends to an antibody molecule such as a polyclonal or monoclonal antibody or an immunologically interactive part or fragment thereof which is capable of binding to a cellulose gene product according to any of the foregoing embodiments.

The term "antibody" as used herein, is intended to include fragments thereof which are also specifically reactive with a polypeptide of the invention. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as for whole antibodies. For example, F(ab')2 fragments can be generated by treating antibody with pepsin. The resulting F(ab')2 fragment can be treated to reduce disulfide bridges to produce Fab' fragments.

Those skilled in the art will be aware of how to produce antibody molecules when provided with the cellulose gene product of the present invention. For example, by using a polypeptide of the present invention polyclonal antisera or monoclonal antibodies can be made using standard methods. A mammal, (e.g., a mouse, hamster, or rabbit) can be immunized with an immunogenic form of the polypeptide which elicits an antibody response in the mammal. Techniques for conferring immunogenicity on a polypeptide include conjugation to carriers or other techniques well known in the art. For example, the polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassay can be used with the immunogen as antigen to assess the levels of antibodies. Following immunization, antisera can be obtained and, if desired IgG molecules corresponding to the polyclonal antibodies may be isolated from the sera.

To produce monoclonal antibodies, antibody producing cells (lymphocytes) can be harvested from an immunized animal and fused with myeloma cells by standard somatic cell fusion

procedures thus immortalizing these cells and yielding hybridoma cells. Such techniques are well known in the art. For example, the hybridoma technique originally developed by Kohler and Milstein (1975) as well as other techniques such as the human B-cell hybridoma technique (Kozbor *et al.*, 1983), the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.*, 1985), and screening of combinatorial antibody libraries (Huse *et al.*, 1989). Hybridoma cells can be screened immunochemically for production of antibodies which are specifically reactive with the polypeptide and monoclonal antibodies isolated.

As with all immunogenic compositions for eliciting antibodies, the immunogenically effective amounts of the polypeptides of the invention must be determined empirically. Factors to be considered include the immunogenicity of the native polypeptide, whether or not the polypeptide will be complexed with or covalently attached to an adjuvant or carrier protein or other carrier and route of administration for the composition, i.e. intravenous, intramuscular, subcutaneous, *etc.*, and the number of immunizing doses to be administered. Such factors are known in the vaccine art and it is well within the skill of immunologists to make such determinations without undue experimentation.

It is within the scope of this invention to include any second antibodies (monoclonal, polyclonal or fragments of antibodies) directed to the first mentioned antibodies discussed above. Both the first and second antibodies may be used in detection assays or a first antibody may be used with a commercially available anti-immunoglobulin antibody.

The present invention is further described by reference to the following non-limiting Figures and Examples.

In the Figures:

Figure 1 is a photographic representation showing the inflorescence length of wild-type *Arabidopsis thaliana* Columbia plants (plants 1 and 3) and *rsw*1 plants (plants 2 and 4) grown at 21°C (plants 1 and 2) or 31°C. Plants were grown initially at 21°C until bolting commenced, the bolts were removed and the re-growth followed in plants grown at each temperature.

Figure 2 is a photographic representation of a cryo-scanning electron micrograph showing misshapen epidermal cells in the cotyledons and hypocotyl of the *rsw*1 mutant when grown at 31°C for 10 days.

Figure 3 is a graphical representation of a gas chromatograph of alditol acetates of methylated sugars from a cellulose standard (top panel) and from the neutral glucan derived from shoots of *rsw*1 plants grown at 31°C (lower panel). The co-incident peaks show that the *rsw*1 glucan is 1,4-linked.

Figure 4 is a schematic representation of the contiguous region of *Arabidopsis thaliana* chromosome 4 (stippled box) between the cosmid markers g8300 and 06455, showing the location of overlapping YAC clones (open boxes) within the contiguous region. The position of the *RSW*1 locus is also indicated, approximately 1.2cM from g8300 and 0.9cM from 06455. The scale indicates 100kb in length. L, left-end of YAC; R, right-end of YAC. Above the representation of chromosome 4, the YAC fragments and cosmid clone fragments used to construct the contiguous region are indicated, using a prefix designation corresponding to the YAC or cosmid from which the fragments were obtained(eg yUP9E3, yUP20B12, etc) and a suffix designation indicating whether the fragment corresponds to the right-end (RE) or left-end (LE) of the YAC clone; N, North; S, South; CAPS, cleaved amplified polymorphic sequence (Konieczny and Ausubel, 1993) version of the g8300 marker.

Figure 5 is a schematic representation of a restriction map of construct 23H12 between the left T-DNA border (LB) and right T-DNA border (RB) sequences (top solid line), showing the position of the *Arabidopsis thaliana RSW1* locus (stippled box). The line at the top of the figure indicates the region of 23H12 which is contained in construct pRSW1. The structure of the *RSW1* gene between the translation start (ATG) and translation stop (TAG) codons is indicated at the bottom of the figure. Exons are indicated by filled boxes; introns are indicated by the solid black line. The alignment of EST clone T20782 to the 3'-end of the *RSW1* gene, from near the end of exon 7 to the end of exon 14, is also indicated at the bottom of the figure. Restriction sites within 23H12 are as follows: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I; Sm, *Sma*I.

Figure 6 is a photographic representation showing complementation of the radial root swelling phenotype of the *rsw1* mutant by transformation with construct 23H12. The *rsw1* mutant was transformed with 23H12 as described in Example 6. Transformed *rsw1* plants (centre group of three seedlings), untransformed *rsw1* plants (left group of three seedlings) and untransformed *A.thaliana* Columbia plants (right group of three seedlings) were grown at 21°C for 5 days and then transferred to 31°C for a further 2 days, after which time the degree of root elongation and radial root swelling was determined.

Figure 7 is a photographic representation comparing wild-type *Arabidopsis thaliana* Columbia plants (right-hand side of the ruler) and *A.thaliana* Columbia plants transformed with the antisense *RSW1* construct (i.e. EST T20782 expressed in the antisense orientation under control of the CaMV 35S promoter sequence; left-hand side of the ruler), showing inflorescence shortening at 21°C in plants transformed with the antisense *RSW1* construct compared to untransformed Columbia plants. The phenotype of the antisense plants at 21°C is similar to the phenotype of the *rsw1* mutant at 31°C. Inflorescence height is indicated in millimetres.

Figure 8 is a schematic representation showing the first 90 amino acid residues of *Arabidopsis thaliana* RSW1 aligned to the amino acid sequences of homologous polypeptides

from *A. thaliana* and other plant species. The shaded region indicates highly conserved sequences. Ath-A and Ath-B are closely related *Arabidopsis thaliana* cDNA clones identified by hybridisation screening using part of the *RSW1* cDNA as a probe. S0542, rice EST clone (MAFF DNA bank, Japan); *cel*A1 and *cel*A2, cotton cDNA sequences expressed in cotton fibre (Pear *et al*, 1996); SOYSTF1A and SOYSTF1B, putative soybean bZIP transcription factors. Amino acid designations are as indicated in Table 1 incorporated herein. Conserved cysteine residues are indicated by the asterisk.

Figure 9 is a schematic representation showing the alignment of the complete amino acid sequence of *Arabidopsis thaliana* RSW1 to the amino acid sequences of homologous polypeptides from *A. thaliana* and other plant species. The shaded region indicates highly conserved sequences. Ath-A and Ath-B are closely related *Arabidopsis thaliana* cDNA clones identified by hybridisation screening using part of the *RSW1* cDNA as a probe. S0542, rice EST clone (MAFF DNA bank, Japan); *cel*A1, cotton genetic sequence (Pear *et al*, 1996); D48636, a partial cDNA clone obtained from rice (Pear *et al*, 1996). Amino acid designations are as indicated in Table 1 incorporated herein. Numbering indicates the amino acid position in the RSW1 sequence.

Figure 10 is a schematic representation of the RSW1 polypeptide, showing the positions of putative transmembrane helices (hatched boxes), cysteine-rich region (Cys) and aspartate residues (D) and the QVLRW signature which are conserved between RSW1 and related amino acid sequences. Regions of RSW1 which are highly-conserved between putative cellulose biosynthesis polypeptides are indicated by the dark-shaded boxes, while less-conserved regions are indicated by the light-shaded boxes.

Figure 11 is a photographic representation of a Southern blot hybridisation of the 5'- end of the *Arabidopsis thaliana* RSW1 cDNA to *Bgl*II-digested DNA derived from *A. thaliana* (lane 1) and cotton (lane 2). Hybridisations were carried out under low stringency conditions at 55°C. Arrows indicate the positions of hybridising bands.

EXAMPLE 1

CHARACTERISATION OF THE CELLULOSE-DEFICIENT *Arabidopsis thaliana* MUTANT *rsw1*

1. Morphology

The *Arabidopsis thaliana rsw1* mutant was produced in a genetic background comprising the ecotype Columbia.

The altered root cell-shape and temperature sensitivity of the root morphology of the *Arabidopsis thaliana* mutant *rsw1* are disclosed, among other morphological mutants, by Baskin *et al.* (1992).

As shown in Figure 1, the present inventors have shown that the *rsw1* mutant exhibits the surprising phenotype of having reduced inflorescence height when grown at 31°C, compared to wild-type Columbia plants grown under similar conditions. In contrast, when grown at 21°C, the inflorescence height of *rsw1* is not significantly different from wild type plants grown under similar conditions, indicating that the shoot phenotype of *rsw1* is conditional and temperature-dependent.

Furthermore, cryo-scanning electron microscopy of the epidermal cells of the *rsw1* mutant indicates significant abnormality in cell shape, particularly in respect of those epidermal cells forming the leaves, hypocotyl and cotyledons, when the seedlings are grown at 31°C (Figure 2).

Rosettes (terminal complexes) are the putative hexameric cellulose synthase complexes of higher plant plasma membranes (Herth, 1985). Freeze-fractured root cells of *Arabidopsis thaliana rsw1* plants grown at 18°C show cellulose microfibrils and rosettes on the PF face of the plasma membrane that resembles those of wild-type *A. thaliana* and other angiosperms. Transferring the *rsw1* mutant to 31°C reduces the number of rosettes in the mutant within 30 min, leading to extensive loss after 3 hours. Plasma membrane particles

align in rows on prolonged exposure to the restrictive temperature. In contrast, there is no change in the appearance of cortical microtubules that align cellulose microfibrils, or of Golgi bodies that synthesise other wall polysaccharides and assemble rosettes.

2. Carbohydrate content

The effect of mutations in the *RSW1* gene on the synthesis of cellulose and other carbohydrates was assessed by measuring *in vivo* incorporation of ^{14}C (supplied as uniformly labelled glucose) into various cell wall fractions. Wild type (*RSW1*) and homozygous mutant *rsw1* seed were germinated at 21°C on agar containing Hoagland's nutrients and 1% (w/v) unlabelled glucose. After 5 d, half of the seedlings were transferred to 31°C for 1 d while the remainder were maintained at 21°C for the same time. Seedlings were covered with a solution containing Hoagland's nutrients and ^{14}C-glucose and incubated for a further 3 h at the same temperature. Rinsed roots and shoots were separated and frozen in liquid nitrogen. Tissue was homogenised in cold, 0.5 M potassium phosphate buffer (0.5M KH_2PO_4, pH7.0) and a crude cell wall fraction collected by centrifugation at 2800 rpm. The wall fraction was extracted with chloroform/methanol [1:1 (v/v)] at 40°C for 1 hour, followed by a brief incubation at 150°C, to remove lipids. The pellet was washed successively with 2ml methanol, 2ml acetone and twice with 2ml of deionised water. Finally, the pellet was extracted successively with dimethyl sulphoxide under nitrogen to remove starch; 0.5% ammonium oxalate to remove pectins; 0.1 M KOH and 3 mg/ml $NaBH_4$ and then with 4 M KOH and 3 mg/ml $NaBH_4$ to extract hemicelluloses; boiling acetic acid/nitric acid/water [8:1:2 (v/v)], to extract any residual non-cellulosic carbohydrates and leave crystalline cellulose as the final insoluble pellet (Updegraph, 1969). All fractions were analysed by liquid scintillation counting and the counts in each fraction from the mutant were expressed as a percentage of the counts in the wild type under the same conditions.

As shown in Table 3, mutant and wild type plants behave in quite similar fashion at 21°C (the permissive temperature) whereas, at the restrictive temperature of 31°C, the incorporation of ^{14}C into cellulose is severely inhibited (to 36% of wild type) by the *rsw1* mutation. The data in Table 3 indicate that cellulose synthesis is specifically inhibited in

the *rsw*1 mutant. The wild type *RSW*1 gene is therefore involved quite directly in cellulose synthesis and changing its sequence by mutation changes the rate of synthesis.

TABLE 3

Counts in fractions from *rsw*1 plants expressed as a % of counts in comparable fraction from wild type plants					
Pectins		Hemicelluloses		Cellulose	
21°C	31°C	21°C	31°C	21°C	31°C
125	104	111	101	80	36

In homozygous mutant *rsw*1 plants, the pectin fraction extracted by ammonium oxalate contained abundant glucose, atypical of true uronic acid-rich pectins. The great majority of the glucose remained in the supernatant when cetyltrimethylammonium bromide precipitated the negatively charged pectins.

3. Non-crystalline β-1,4-glucan content

The quantity of cellulose and the quantity of a non-crystalline β-1,4-glucan recovered from the ammonium oxalate fraction were determined for seedlings of wild type Columbia and for backcrossed, homozygous *rsw*1 that were grown for either 7 days at 21°C or alternatively, for 2 days at 21°C and 5 days at 31°C, on vertical agar plates containing growth medium (Baskin *et al.*, 1992) plus 1% (w/v) glucose, and under continuous light (90 μmol m^{-2} s^{-1}). Roots and shoots were separated from about 150 seedlings, freeze-dried to constant weight and ground in a mortar and pestle with 3 ml of cold 0.5 M potassium phosphate buffer (pH 7.0). The combined homogenate after two buffer rinses (2ml each) was centrifuged at 2800 x g for 10 min. After washing the pellet fraction twice with 2 ml buffer and twice with 2 ml distilled water, the pellet, comprising the crude cell wall fraction, and the pooled supernatants, comprising the phosphate buffer fraction were retained. The crude cell wall pellet fraction was stirred with two 3 ml aliquots of chloroform/methanol [1:1 (v/v)] for 1 hour at 40°C, 2 ml of methanol at 40°C for 30 min, 2 ml of acetone for 30 min, and twice with water. The whole

procedure repeated in the case of shoots. Combined supernatants were dried in a nitrogen stream. The pellet was successively extracted with: (i)3 ml of DMSO- water 9:1 [v/v], sealed under nitrogen, overnight with shaking, followed by two 2ml extractions using DMSO/water and three 2ml water washes; (ii) 3ml of ammonium oxalate (0.5 %) at 100°C for 1 hour, followed by two water washes; (iii) 3ml of 0.1 M KOH containing 1mg/ ml sodium borohydride, for 1 hour at 25°C (repeated once for root material or twice for shoot material), with a final wash with 2 ml water; (iv) 3 ml of 4 M KOH containing 1 mg/ml sodium borohydride, for 1 hour at 25°C (repeated once for root material or twice for shoot material). The final pellet was boiled with intermittent stirring in 3 ml of acetic acid-nitric acid-water [8:1:2 (v/v)] (Updegraph 1969), combined with 2 water washes, and diluted with 5 ml water.

The insoluble residue of cellulose was solubilised in 67% (v/v) H_2SO_4, shown to contain greater than 97% (w/v) glucose using GC/MS (Fisons AS800/MD800) of alditol acetates (Doares *et al.*, 1991) and quantified in three independent samples by anthrone/H_2SO_4 reaction. Results of GC/MS for pooled replica samples are presented in Table 4.

The non-crystalline β-1,4-glucan was recovered as the supernatant from the ammonium oxalate fraction when anionic pectins were precipitated by overnight incubation at 37°C with 2% (w/v) cetyltrimethylammonium bromide (CTAB) and collected by centrifugation at 2800 x g for 10 min. The glucan (250 μg/ml) or starch (Sigma; 200 μg/ml) were digested with mixtures of endocellulase (EC 3.2.1.4; Megazyme, Australia) from *Trichoderma* and almond β-glucosidase (EC 3.2.1.21; Sigma), or *Bacillus sp.* α-amylase (EC 3.2.1.1; Sigma) and rice α-glucosidase (EC 3.2.1.20; Sigma).

The material recovered in the supernatant from the ammonium oxalate fraction was shown to contain a pure β-1,4-glucan by demonstrating that: (i) only glucose was detectable when it was hydrolysed by 2 M TFA in a sealed tube for 1 h at 120°C in an autoclave, the supernatant (2000 g for 5 min) was dried under vacuum at 45°C to remove TFA and glucose was determined by GC/MS; (ii) methylation (Needs and Selvendran 1993) gave a dominant peak resolved by thin layer chromatography and by GC/MS

that was identical to that from a cellulose standard and so indicative of 1,4-linked glucan (Figure 3); and

(iii) the endo-cellulase and β-1,4-glucosidase mixture released 83 % of the TFA-releasable glucose from the glucan produced by *rsw*1 at 31 °C while the α-amylase/α-glucosidase mixture released no glucose from the glucan. Conversely, the α-amylase/ α-glucosidase mixture released 95% of the TFA-releasable glucose from a starch sample, while the endo-cellulase/β-1,4-glucosidase mixture released no glucose from starch.

Extractability of the glucan using ammonium oxalate, and the susceptibility of the glucan to endocellulase/β-glucosidase and TFA hydrolysis indicate that the glucan in the *rsw*1 mutant is not crystalline, because it is the crystallinity of glucan which makes cellulose resistant to extraction and degradation.

Table 4 shows the quantity of glucose in cellulose determined by the anthrone/H_2SO_4 reaction and the quantity in the non-crystalline glucan after TFA hydrolysis, for shoots of wild type and mutant *rsw*1 *Arabidopsis* plants. The data indicate that the production of cellulose and of the non-crystalline β-1,4-glucan can be manipulated by mutational changes in the RSW1 gene.

TABLE 4

Glucose contents of cellulose and of the ammonium oxalate-extractable glucan

	wild type		*rsw*1	
	21°C	**31°C**	**21°C**	**31°C**
Cellulose	273+28	363+18*	218+20	159+19*
Glucan	22	58	24	195

All values nmol glucose mg-1 plant dry weight + sd (n=3).

* Differences significant at 0.001 % level.

4. Starch content

The quantity of starch recovered in the DMSO fraction from roots in the experiment described above was also determined by the anthrone/H_2SO_4 extraction (Table 5).

As shown in Table 5, the level of starch deposited in the *rsw*1 mutant is 4-fold that detectable in the roots of wild-type plants at the restrictive temperature of 31°C. A similar rise in starch is also seen if the data are expressed as nmol glucose per plant. There is no detectable difference in deposition at starch between *rsw*1 plants and wild-type plants at 21°C.

TABLE 5

Quantity of starch (nmol glucose per mg dry weight of seedling) extracted from roots of *rsw*1 and wild type seedlings

Temperature	Phenotype	
	Wild-type	*rsw*1 mutant
21°C	22	18
31°C	37	126

The composition of cell walls in the *rsw*1 mutant plant compared to wild type plants at the restrictive temperature of 31°C, is summarised in Table 6.

TABLE 6

Mol% composition of cell walls from shoots of *rsw*1 and wild-type seedlings grown at 31°C

Cell wall component	Phenotype	
	Wild-type	*rsw*1 mutant
Crystalline cellulose	38.4	16.5
Non- crystalline β-1,4-glucan	8.5	27.1
Pectin	37.1	36.3
Alkali-soluble	15.6	19.8
Acid-soluble	0.3	0.4

In conclusion, the *rsw*1 mutation disassembles cellulose synthase complexes in the plasma membrane, reduces cellulose accumulation and causes β-1,4-glucan to accumulate in a non-crystalline form.

EXAMPLE 2

MAPPING OF YAC CLONES TO THE *rsw*1 LOCUS

The *rsw*1 locus in the mutant *Arabidopsis thaliana* plant described in Example 1 above was mapped to chromosome 4 of *A. thaliana* using RFLP gene mapping techniques(Chang *et al*, 1988; Nam *et al.*, 1989) to analyse the F_2 or F_3 progeny derived from a Columbia (Co)/Landsberg (Ler) cross. In particular, the *rsw*1 mutation was shown to be linked genetically to the *ga*5 locus, which is a chromosome 4 visual marker in *A. thaliana*.

Based on an analysis of map distances and chromosomal break points in 293 F_2 or F_3 progeny derived from a Columbia (Co)/Landsberg (Ler) cross, *rsw*1 was localised to an approximately 2.1 cM region between the RFLP markers g8300 and 06455, approximately 1.2cM south of the CAPS (cleaved amplified polymorphic sequence; Konieczny and Ausubel, 1993) version of the g8300 marker (Figure 4).

The interval between g8300 and 06455 in which *rsw*1 residues was found to be spanned by an overlapping set of Yeast Artificial Chromosome (YAC) clones. The clones were obtained from Plant Industry, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia. The YACs were positioned in the g8300/06455 interval by hybridisation using known DNA molecular markers (from within the interval) and DNA fragments from the ends of the YACs. The length of the interval was estimated to comprise 900kb of DNA.

Refined gene mapping of recombinants within the region spanned by YAC clones established the genetic distance between the RFLP marker g8300 and the *rsw*1 locus.

The combination of genetic map distance data and the mapping of YAC clones within the region further localised the *rsw*1 locus to the YAC clone designated yUP5C8.

EXAMPLE 3

MAPPING OF cDNA CLONES TO THE YAC CLONE YUP5C8

An *Arabidopsis thaliana* cDNA clone designated T20782 was obtained from the public *Arabidopsis* Resource Centre, Ohio State University, 1735 Neil Avenue, Columbus, OH 43210, United States of America. The T20782 cDNA clone was localised broadly to the DNA interval on *Arabidopsis* chromosome 4 between the two markers g8300 and 06455 shown in Figure 4. Using a polymerase chain reaction (PCR) based approach DNA primers (5'-AGAACAGCAGATACACGGA-3' and 5'-CTGAAGAAGGCTGGACAAT-3') designed to the T20782 cDNA nucleotide sequence were used to screen *Arabidopsis* YAC clone libraries. The T20782 cDNA clone was found to localise to YACs (CIC1F9, CIC10E9, CIC11D9) identified on the *Arabidopsis* chromosome 4 g8300 and 06455 interval (Figure 4). The same approach was used to further localise clone T20782 to YAC clone yUP5C8, the same YAC designated to contain the *rsw*1 locus in the same chromosome interval (Figure 4).

Furthermore, amplification of the YAC clone yUP5C8 using primers derived from T20782 produces a 500bp fragment containing two putative exons identical to part of the T20782 nucleotide sequence, in addition to two intron sequences.

The cDNA T20782 was considered as a candidate gene involved in cellulose biosynthesis.

EXAMPLE 4

NUCLEOTIDE SEQUENCE ANALYSIS OF THE CDNA CLONE T20782

The nucleotide sequence of the cDNA clone T20782 is presented in SEQ ID NO: 1. The nucleotide sequence was obtained using a Dye Terminator Cycle Sequencing kit (Perkin Elmer cat. #401384) as recommended by the manufacturer. Four template clones were used for nucleotide sequencing to generate the sequence listed. The first template was the cDNA clone T20782. This template was sequenced using the following sequencing primers:

a)5'-CAATGCATTCATAGCTCCAGCCT-3'

b)5'-AAAAGGCTGGAGCTATGAATGCAT-3'

c)5'-TCACCGACAGATTCATCATACCCG-3'

d)5'- GACATGGAATCACCTTAACTGCC-3'

e)5'-CCATTCAGTCTTGTCTTCGTAACC-3'

f)5'-GGTTACGAAGACAAGACTGAAATGG-3'

g)5'-GAACCTCATAGGCATTGTGGGCTGG-3'

h)5'-GCAGGCTCTATATGGGTATGATCC-3'

i)Standard M13 forward sequencing primer.

j)Standard T7 sequencing primer.

The second template clone (T20782 *Sph*I deletion clone) was constructed by creating a DNA deletion within the T20782 clone. The T20782 clone was digested with the restriction enzyme *Sph*I, the enzyme was heat-killed, the DNA ligated and electroporated into NM522 *E.coli* host cells. The T20782 *Sph*I deletion clone was then sequenced using a standard M13 forward sequencing primer. Two other deletion clones were made for DNA sequencing in a similar fashion but the restriction enzymes *Eco*RI and *Sma*I were used. The T20782 *Eco*RI deletion clone and the T20782 *Sma*I deletion clone were sequenced using a standard T7 sequencing primer. The DNA sequence shown in SEQ ID NO:1 is for one DNA strand only however those skilled in the art will be able to generate the nucleotide sequence of the

complementary strand from the data provided.

The amino acid sequence encoded by clone T20782 was derived and is set forth in SEQ ID NO:2.

The T20782 clone encodes all but the first Aspartate (D) residue of the D, D, D, QXXRW signature conserved in the general architecture of β-glycosyl transferases. In particular, T20782 encodes 5 amino acid residues of the D, D, D, QXXRW signature, between amino acid positions 109 and 370 of SEQ ID NO:2. The conserved Aspartate, Aspartate, Glutamine, Arginine and Tryptophan amino acid residues are shown below, in bold type, with the local amino acid residues also indicated:

1. Amino acid residues 105 to 113 of SEQ ID NO:2:
   LLNV**D**C**D**H**Y**;
2. Amino acid residues 324 to 332 of SEQ ID NO:2:
   SVTE**D**ILT**G**; and
3. Amino acid residues 362 to 374 of SEQ ID NO:2:
   DRLN**Q**VLR**W**ALGS.

It must be noted that these invariable amino acids merely indicate that the T20782 derived amino acid sequence belongs to a very broad group of glycosyl transferases. Some of these enzymes such as cellulose synthase, chitin synthase, alginate synthase and hyaluronic acid synthase produce functionally very different compounds.

The presence of the conserved amino acid residues merely indicate that the T20782 clone may encode a β-glycosyl transferase protein such as the cellulose gene product, cellulose synthase. The fact that the clone localises in the vicinity of a gene involved in cellulose biosynthesis is the key feature which now focus interest on the T20782 clone as a candidate for the *RSW1* (cellulose synthase) gene.

The T20782 potentially codes for a cellulose synthase.

EXAMPLE 5

NUCLEOTIDE SEQUENCE ANALYSIS OF THE GENOMIC CLONE 23H12

Clone 23H12 contains approximately 21kb of *Arabidopsis thaliana* genomic DNA in the region between the left border and right border T-DNA sequences, and localises to the *RSW1* candidate YAC yUP5C8. Clone 23H12 was isolated by hybridisation using EST20782 insert DNA, from a genomic DNA library made for plant transformation. Cosmid 12C4 was also shown to hybridize to the cDNA clone T20782, however this cosmid appears to comprise a partial genomic sequence corresponding to the related *Ath*-A cDNA sequence set forth in SEQ ID NO:7, for which the corresponding amino acid sequence is set forth in SEQ ID NO:8.

A restriction enzyme map of clone 23H12 is presented in Figure 5.

Nucleotide sequence of 8411bp of genomic DNA in the binary cosmid clone 23H12 was obtained (SEQ ID NO:3) by primer walking along the 23H12 template, using a Dye Terminator Cycle Sequencing kit (Perkin Elmer cat. #401384) as recommended by the manufacturer. The following primers at least, were used for DNA sequencing of the 23H12 clone DNA:

a)cs1-R	5'-CAATGCATTCATAGCTCCAGCCT-3'
b)cs1-F	5'-AAAAGGCTGGAGCTATGAATGCAT-3'
c)up	5'-AGAACAGCAGATACACGGA-3'
d)ve76-R2	5'-ATCCGTGTATCTGCTGTTCTTACC-3'
e)est1-R	5'-AATGCTCTTGTTGCCAAAGCAC-3'
f)sve76-F	5'-ATTGTCCAGCCTTCTTCAGG-3'
g)ve76-R	5'-CTGAAGAAGGCTGGACAATGC-3'

h)B12-R1 5'-AGGTAAGCATAGCTGAACCATC-3'

i)B12-R2 5'-AGTAGATTGCAGATGGTTTTCTAC-3'

j)B12-R3 5'-TTCAATGGGTCCACTGTACTAAC-3'

k)B12-R4 5'-ATTCAGATGCACCATTGTC-3'

The structure of the *RSW1* gene contained in cosmid clone 23H12 is also presented in Figure 5. As shown therein, coding sequences in 23H12, from the last 12 bp of exon 7 to the end of exon 14, correspond to the full T20782 cDNA sequence (i.e. SEQ ID NO:1). The nucleotide sequences of the *RSW1* gene comprising exons 1 to 8 were amplified from *A.thaliana* Columbia double-stranded cDNA, using amplification primers upstream of the *RSW1* start site and a primer internal to the EST clone T20782.

The exons in the *RSW1* gene range from 81bp to 585bp in length and all 5' and 3' intron/exon splice junctions conform to the conserved intron rule.

The *RSW1* transcript comprises a 5'-untranslated sequence of at least 70bp in length, a 3243bp coding region and a 360bp 3'-untranslated region. Northern hybridization analyses indicate that the *RSW1* transcript in wild-type *A. thaliana* roots, leaves and inflorescences is approximately 4.0kb in length, and that a similar transcript size occurs in mutant tissue (data not shown).

The derived amino acid sequence of the RSW1 polypeptide encoded by the cosmid clone 23H12 (i.e. the polypeptide set forth in SEQ ID NO:6) is 1081 amino acids in length and contains the entire D, D, D, QXXRW signature characteristic of β-glycosyl transferase proteins, between amino acid position 395 and amino acid position 822. The conserved Aspartate, Glutamine, Arginine and Tryptophan residues are shown below, in bold type, with the local amino acid residues also indicated:

1. amino acid residues 391 to 399 of SEQ ID NO:6:

YVSDDGSAM

2. Amino acid residues 557 to 565 of SEQ ID NO:6:
LLNVDCDHY;
3. Amino acid residues 776 to 784 of SEQ ID NO:6:
SVTEDILTG; and
4. Amino acid residues 814 to 826 of SEQ ID NO:6:
DRLNQVLRWALGS.

The second and third conserved Aspartate residues listed *supra*, and the fourth conserved amino acid sequence motif listed *supra* (i.e. QVLRW) are also present in the cDNA clone T20782 (see Example 4 above).

The 23H12 clone potentially encodes a cellulose synthase.

EXAMPLE 6

COMPLEMENTATION OF THE *rsw1* MUTATION

The complementation of the cellulose mutant plant *rsw1* is the key test to demonstrate the function of the clone 23H12 gene product. Complementation of the *rsw1* phenotype was demonstrated by transforming the binary cosmid clone 23H12, or a derivative clone thereof encoding a functional gene product, into the *Arabidopsis thaliana* cellulose mutant *rsw1*. Two DNA constructs (23H12 and pRSW1) were used to complement the *rsw1* mutant plant line.

1. Construct 23H12

Clone 23H12 is described in Example 5 and Figure 5.

2. Construct pRSW1

The 23H12 construct has an insert of about 21kb in length. To demonstrate that any complementation of the phenotype of the *rsw1* mutation is the result of expression of the gene

which corresponds to SEQ ID NO:3, a genetic construct, designated as pRSW1, comprising the putative RSW1 gene with most of the surrounding DNA deleted, was produced. A restriction enzyme (RE) map of the *RSW*1 gene insert in pRSW1 is provided in Figure 5.

To produce pRSW1, the RSW1 gene was subcloned from cosmid 23H12 and cloned into the binary plasmid pBIN19. Briefly, *Escherichia coli* cells containing cosmid 23H12 were grown in LB medium supplemented with tetracyclin (3.5 mg/L). Plasmid DNA was prepared by alkaline lysis and digested sequentially with restriction enzymes *Pvu*II and *Sal*I. Two co-migrating fragments of 9 kb and 10 kb, respectively, were isolated as a single fraction from a 0.8% (w/v) agarose gel. The RSW1 gene was contained on the 10 kb *Pvu*II/*Sal*I fragment. The 9 kb fragment appeared to be a *Pvu*II cleavage product not comprising the RSWl gene. The restriction fragments were ligated into pBIN19 digested with *Sma*I and *Sal*I. An aliquot of the ligation mix was introduced by electroporation into *E.coli* strain XLB1. Colonies resistant to kanamycin (50 mg/L) were selected and subsequently characterised by restriction enzyme analysis to identify those clones which contained only the 10 kb *Pvu*II/*Sal*I fragment comprising the RSW1 gene, in pBIN19.

3. Transfer of the 23HI2 and pRSW1 constructs to *Agrobacterium tumefaciens*

Cosmid 23H12 was transferred to Agrobacterium by triparental mating, essentially as described by Ditta *et al.* (1980). Three bacterial strains as follows were mixed on solid LB medium without antibiotics: Strain 1 was an *E. coli* helper strain containing the mobilising plasmid pRK2013, grown to stationary phase; Strain 2 was *E.coli* containing cosmid 23H12, grown to stationary phase; and Strain 3 was an exponential-phase culture of *A. tumefaciens* strain AGL1 (Lazo *et al.*, 1991). The mixture was allowed to grow over night at 28°C, before an aliquot was streaked out on solid LB medium containing antibiotics (ampicillin 50 mg/L, rifampicin 50 mg/L, tetracyclin 3.5 mg/L) to select for transformed *A. tumefaciens* AGL1. Resistant colonies appeared after 2-3 days at 28°C and were streaked out once again on selective medium for further purification. Selected colonies were then subcultured in liquid LB medium supplemented with rifampicin (50 mg/L) and tetracyclin (3.5 mg/L) and stored at -80°C.

Plasmid pRSW1 (initially designated as p2029) was introduced into *A. tumefaciens* strain AGL1 by electroporation.

4. Transformation of *rsw*1 plants

The *rsw*l plant line was transformed with constructs 23H12 and pRSW1 using vacuum infiltration essentially as described by Bechtold et al. (1993).

5. Analysis of radial swelling in transformants

Complementation of the radial swelling (rsw) phenotype , which is characteristic of the *rsw*1 mutant plant, was assayed by germinating transformed (i.e. T1 seed) *rsw*1 seeds obtained as described *supra* on Hoaglands plates containing 50μg/ml kanamycin. Plates containing the transformed seeds were incubated at 21 °C for 10-12 days. Kanamycin-resistant seedlings were transferred to fresh Hoaglands plates containing 50μg/ml kanamycin and incubated at 31 °C for 2 days . Following this incubation, the root tip was examined for a radial swelling phenotype. Under these conditions, the roots of wild-type plants do not show any radial swelling phenotype however, the roots of *rsw*1 plants show clear radial swelling at the root tip and also have a short root compared to the wild-type plants. As a consequence, determination of the radial swelling phenotype of the transformed plants was indicative of successful complementation of the *rsw*1 phenotype.

The kanamycin-resistant seedlings were maintained by further growth of seedlings at 21 °C, following the high temperature incubation. Once plants had recovered, the seedlings were transferred to soil and grown in cabinets at 21 °C (16 hr light/8 hr dark cycle). T2 seed was then harvested from mature individual plants.

Using the 23H12 construct for *rsw*l transformation, a total of 262 kanamycin-resistant seedlings were obtained. All of these transformants were tested for complementation of the root radial swelling phenotype. A total of 230 seedlings showed a wild type root phenotype, while only 32 seedlings showed the radial swelling root phenotype characteristic of *rsw*1 plants. By way of example, Figure 6 shows the phenotypes of transformed seedlings compared

to untransformed wild-type and *rsw*1 seedlings, following incubation at 31°C. As shown in Figure 6, there is clear complementation of the radial swelling phenotype in the transformed seedlings, with normal root length being exhibited by the transformed seedlings at 31°C

Using the pRSW1 construct for transformation, a total of 140 kanamycin-resistant seedlings were obtained. All of the 11 seedlings tested for complementation of the root radial swelling phenotype showed a wild type root phenotype and none of the seedlings showed any signs of radial swelling in the roots (data not shown).

6. General morphological analysis of the complemented *rsw*1 mutant line

Further characterisation of the complemented *rsw*1 plants has shown that other morphological characteristics of *rsw*1 have also been restored in the transgenic lines, for example the bolt (inflorescence) height, and the ability of the plants to grow wild type cotyledons, leaves, trichomes, siliques and flowers at 31°C (data not shown).

7. Biochemical complementation of the *rsw*1 mutant line

T2 seed from transformations using cosmid 23H12 as described *supra* or alternatively, using the binary plasmid pBin19 which lacks any *RSW*1 gene sequences, was sown on Hoagland's solid media containing kanamycin (50μg/ml), incubated for 2 days at 21°C and then transferred to 31°C for 5 days. Wild-type *A.thaliana* Columbia plants were grown under similar conditions but without kanamycin in the growth medium. Kanamycin resistant T2 seedlings which have at least one copy of the 23H12 cosmid sequence, and wild-type seedlings, were collected and frozen for cellulose analysis.

Cellulose levels were determined as acetic-nitric acid insoluble material (Updegraph, 1969) for 10 lines of kanamycin-resistant T2 plants transformed with the 23H12 cosmid sequence, and compared to the cellulose levels in *rsw*1 mutant plants, wild-type *A.thaliana* Columbia plants and *A.thaliana* Columbia plants transformed with the binary plasmid pBin19. The results are provided in Table 7.

As shown in Table 7, the cellulose levels have been significantly elevated in the complemented *rsw*1 (T2) plants, compared to the cellulose levels measured in the *rsw*1 mutant parent plant. In fact, cellulose levels in the 23H12-transformed plants, expressed relative to the fresh weight of plant material or on a per seedling basis, are not significantly different from the cellulose levels of either wild-type *Arabidopsis thaliana* Columbia plants or *A.thaliana* Columbia transformed with the binary plasmid pBin19. These data indicate that the 23H12 cosmid is able to fully complement the cellulose-deficient phenotype of the *rsw*1 mutant.

Homozygous T3 lines are generated to confirm the data presented in Table 7.

Furthermore, data presented in Table 7 indicate that there is no difference in the rate of growth of the T2 transformed *rsw*1 plants and wild-type plants at 31°C, because the fresh weight of such plants does not differ significantly. In contrast, the fresh weight of mutant *rsw*1 seedlings grown under identical conditions is only approximately 55% of the level observed in T2 lines transformed with 23H12 (range about 30% to about 80%). These data support the conclusion that cellulose levels have been manipulated in the complemented *rsw*1 (T2) plants.

Furthermore, the rate of cellulose synthesis in 23H12-transformed plants and wild-type plants at 31°C, as measured by ^{14}C incorporation is also determined.

Furthermore, the β-1,4-glucan levels and starch levels in the 23H12 transformant lines are shown to be similar to the β-1,4-glucan and starch levels in wild-type plants.

TABLE 7

CELLULOSE LEVELS IN *rsw1* PLANTS TRANSFORMED WITH COSMID CLONE 23H12

PLANT LINE	SAMPLE SIZE (No. of plants)	SEEDLING FRESH WEIGHT (mg)	CELLULOSE (mg cellulose/ 100 mg tissue)	CELLULOSE (mg cellulose/ seedling)
1.2 (*rsw1*+23H12)	126	2.51	1.23	0.031
1.4 (*rsw1*+23H12)	132	2.25	2.50	0.056
2.1 (*rsw1*+23H12)	126	3.23	1.29	0.042
3.1 (*rsw1*+23H12)	127	3.75	1.23	0.046
3.10 (*rsw1*+23H12)	128	3.52	1.69	0.060
4.4 (*rsw1*+23H12)	110	5.14	1.31	0.067
4.5 (*rsw1*+23H12)	125	3.18	1.26	0.040
5.3 (*rsw1*+23H12)	124	2.77	1.17	0.032
9.2 (*rsw1*+23H12)	125	2.26	1.41	0.032
10.8 (*rsw1*+23H12)	126	2.4	1.20	0.029
Columbia/pBin19	106	2.64	1.34	0.035
Columbia	178	2.73	1.18	0.032
rsw1 mutant	179	1.77	0.84	0.015

EXAMPLE 7

DETERMINATION OF THE FULL-LENGTH NUCLEOTIDE SEQUENCE ENCODING THE WILD-TYPE RSW1 POLYPEPTIDE

Arabidopsis thaliana double-stranded cDNA and cDNA libraries were prepared using the CAPFINDER cDNA kit (Clontech). RNA was isolated from wild-type Columbia grown in sterile conditions for 21 days.

Approximately 100,000 cDNA clones in an unamplified cDNA library were screened under standard hybridization conditions at 65°C, using a probe comprising ^{32}P-labelled DNA amplified from double stranded cDNA. To prepare the hybridization probe, the following amplification primers were used:

1. 2280-F:5'GAATCGGCTACGAATTTCCCA 3'
2. 2370-F:5'TTGGTTGCTGGATCCTACCGG 3'
3. csp1-R:5'GGT TCT AAA TCT TCT TCC GTC 3'

wherein the primer combinations were either 2280-F/csp1-R or 2370-F/csp1-R. The primer 2280-F corresponds to nucleotide positions 2226 to 2246 in SEQ ID NO:3, upstream of the translation start site. The primer 2370-F corresponds to nucleotide positions 2314 to 2334 in SEQ ID NO:3, encoding amino acids 7 through 13 of the RSW1 polypeptide. The primer csp1-R comprises nucleotide sequences complementary to nucleotides 588 to 608 of the T20782 clone (SEQ ID NO:1) corresponding to nucleotides 6120 to 6140 of SEQ ID NO:3. The hybridization probes produced are approximately 1858 nucleotides in length (2280-F/csp1-R primer combination) or 1946 nucleotides in length (2370-F/csp1-R primer combination).

Five hybridizing bacteriophage clones were identified, which were plaque-purified to homogeneity during two successive rounds of screening. Plasmids were rescued from the positively-hybridizing bacteriophage clones, using the Stratagene excision protocol for the ZapExpress™ vector according to the manufacturer's instructions. Colony hybridizations

confirmed the identity of the clones.

Isolated cDNA clones were sequenced by primer walking similar to the method described in Examples 4 and 5 *supra*.

A full-length wild-type *RSW1* nucleotide sequence was compiled from the nucleotide sequences of two cDNA clones. First, the 3'-end of the cDNA, encoding amino acids 453-1081 of RSW1, corresponded to the nucleotide sequence of the EST clone T20782 (SEQ ID NO:1). The remaining cDNA sequence, encoding amino acids 1-654 of RSW1, was generated by amplification of the 5'-end from cDNA, using primer 2280-F, which comprises nucleotide sequences approximately 50-70bp upstream of the *RSW1* translation start site in cosmid 23 H12, and primer csp1-R, which comprises nucleotide sequences complementary to nucleotides 588 to 608 of the T20782 clone (SEQ ID NO:1).

Several amplified clones are sequenced to show that no nucleotide errors were introduced by the amplification process. The 5' and 3' nucleotide sequences are spliced together to produce the complete *RSW1* open reading frame and 3'-untranslated region provided in SEQ ID NO:5.

Those skilled in the art will be aware that the 5'-end and 3'-end of the two incomplete cDNAs are spliced together to obtain a full-length cDNA clone, the nucleotide sequence of which is set forth in SEQ ID NO:5.

Of the remaining cDNA clones, no isolated cDNA clone comprised a nucleotide sequence which precisely matched the nucleotide sequence of the *RSW1* gene present in cosmid 23H12. However, several clones containing closely-related sequences were obtained, as summarised in Table 8. The nucleotide sequences of the *Ath*-A and *Ath*-B cDNAs are provided herein as SEQ ID Nos: 7 and 9, respectively.

TABLE 8

CHARACTERISATION OF *A. thaliana* cDNA CLONES

CLONE NAME	DESCRIPTION	LENGTH	SEQ ID NO:
RSW1.1A	chimeric clone	partial	not provided
RSW1A	chimeric clone	partial	not provided
Ath-A	12C4 cDNA	full-length	SEQ ID NO:7
Ath-B	new sequence	full-length	SEQ ID NO:9
RSW4A	identical to *Ath*-B	full-length	not provided

The derived amino acid sequences encoded by the cDNAs listed in Table 8, is provided in Figures 8 and 9 and SEQ ID Nos: 8 and 10 herein.

Figure 10 a schematic representation of the important features of the RSW1 polypeptide which are conserved within *A.thaliana* and between *A.thaliana* and other plant species. In addition to the species indicated in Figure 10, the present inventors have also identified maize, wheat, barley and *Brassica ssp.* cellulose biosynthetic genes by homology search. Accordingly, the present invention extends to cellulose genes and cellulose biosynthetic polypeptides as hereinbefore defined, derived from any plant species, including *A. thaliana*, cotton, rice, wheat, barley, maize, *Eucalyptus ssp.*, *Brassica ssp. Pinus ssp.*, *Populus ssp.*, *Picea ssp.*, hemp, jute and flax, amongst others.

EXAMPLE 8

ISOLATION OF FULL-LENGTH NUCLEOTIDE SEQUENCE ENCODING THE MUTANT RSW1 POLYPEPTIDE

Arabidopsis thaliana double-stranded cDNA and cDNA libraries were prepared using the CAPFINDER cDNA kit (Clontech). RNA was isolated from *Arabidopsis thaliana* Columbia *rsw1* mutant plants grown in sterile conditions for 21 days.

The full-length *rsw1* mutant nucleotide sequence was generated by sequencing two amplified

DNA fragments spanning the *rsw1* mutant gene. The 5'- end sequence of the cDNA (comprising the 5'-untranslated region and exons 1-11) was amplified using the primer combination 2280-F/csp1-R (Example 7). The 3'-end sequence was amplified using the primers EST1-F and cs3-R set forth below:

1.Primer EST1-F: 5'AATGCTTCTTGTTGCCAAAGCA 3'

2.Primer cs3-R: 5'GACATGGAATCACCTTAACTGCC 3'

wherein primer EST1-F corresponds to nucleotide positions 1399-1420 of SEQ ID NO:5 (within exon 8) and primer cs3-R is complementary to nucleotides 3335-3359 of SEQ ID NO:5 (within the 3'-untranslated region of the wild-type transcript).

The full-length sequence of the mutant *rsw1* transcript is set forth herein as SEQ ID NO:11.

Whilst not being bound by any theory or mode of action, a single nucleotide substitution in the *rsw1* mutant nucleotide sequence (nucleotide position 1716 in SEQ ID NO:11), relative to the wild-type *RSW1* nucleotide sequence (nucleotide position 1646 in SEQ ID NO:5), resulting in Ala549 being substituted with Val549 in the mutant polypeptide, may contribute to the altered activity of the RSW1 polypeptide at non-permissive temperatures such as 31°C. Additional amino acid substitutions are also contemplated by the present invention, to alter the activity of the RSW1 polypeptide, or to make the polypeptide temperature-sensitive.

EXAMPLE 9

ANTISENSE INHIBITION OF CELLULOSE PRODUCTION IN TRANSGENIC PLANTS

1. Construction of an antisense *RSW1* binary vector

One example of transgenic plants in which cellulose production is inhibited is provided by the expression of an antisense genetic construct therein. Antisense technology is used to target expression of a cellulose gene(s) to reduce the amount of cellulose produced by

transgenic plants.

By way of exemplification, an antisense plant transformation construct has been engineered to contain the T20782 cDNA insert (or a part thereof) in the antisense orientation and in operable connection with the CaMV 35S promoter present in the binary plasmid pRD410 (Datla *et al*, 1992). More particularly, the T20782 cDNA clone, which comprises the 3'-end of the wild-type *RSW1* gene, was digested with *Xba*I and *Kpn*I and cloned into the kanamycin-resistant derivative of pGEM3zf(-), designated as plasmid, pJKKMf(-). The *RSW1* sequence was sub-cloned, in the antisense orientation, into the binary vector pRD410 as a *Xba*I/*Sac*I fragment, thereby replacing the β-glucuronidase (GUS or *uidA*) gene. This allows the *RSW1* sequence to be transcribed in the antisense orientation under the control of the CaMV 35S promoter.

The antisense *RSW1* binary plasmid vector was transferred to *Agrobacterium tumefaciens* strain AGL1, by triparental mating and selection on rifampicin and kanamycin, as described by Lazo *et al*. (1991). The presence of the *RSW1* insert in transformed *A.tumefaciens* cells was confirmed by Southern hybridization analysis (Southern, 1975). The construct was shown to be free of deletion or rearrangements prior to transformation of plant tissues, by back-transformation into *Escherichia coli* strain JM101 and restriction digestion analysis.

2. Transformation of *Arabidopsis thaliana*

Eight pots, each containing approximately 16 *A. thaliana* ecotype Columbia plants, were grown under standard conditions. Plant tissue was transformed with the antisense *RSW1* binary plasmid by vacuum infiltration as described by Bechtold *et al* (1993). Infiltration media contained 2.5% (w/v) sucrose and plants were infiltrated for 2 min until a vacuum of approximately 400mm Hg was obtained. The vacuum connection was shut off and plants allowed to sit under vacuum for 5 min.

Approximately 34,000 T1 seed was screened on MS plates containing 50μg/ml kanamycin, to select for plants containing the antisense *RSW1* construct. Of the T1 seed sown, 135

kanamycin-resistant seedlings were identified, of which 91 were transferred into soil and grown at 21 °C under a long-day photoperiod (16hr light; 8hr dark).

Of the 91 transgenic lines, 19 lines were chosen for further analysis which had anther filaments in each flower which were too short to deposit pollen upon the stigma and, as a consequence, required hand-pollination to obtain T2 seed therefrom.

T2 seed from 14 of these 19 lines was plated out onto vertical Hoaglands plates containing kanamycin to determine segregation ratios. Between five and ten seed were plated per transgenic line. Control seeds, including *A. thaliana* Columbia containing the binary vector pBIN19 (Bevan, 1984) and segregating 3:1 for kanamycin resistance, and the *rsw*1 mutant transformed with the *NPTII* gene, also segregating 3:1 for kanamycin resistance, were grown under the same conditions. Kanamycin-resistant plants were transferred to soil and grown at 21 °C under long days, until flowering. Untransformed *Arabidopsis thaliana* Columbia plants were also grown under similar conditions, in the absence of kanamycin.

3. Morphology of antisense- *RSW*1 plants

A comparison of the morphology of antisense *RSW*1 plants grown at 21 °C, to mutant *rsw*1 plants grown at the non-permissive temperature (i.e. 31 °C) has identified a number of common phenotypes. For example, the antisense plants exhibit reduced fertility, inflorescence shortening and have short anthers, compared to wild-type plants, when grown at 21 °C. These phenotypes are also observed in mutant *rsw*1 plants grown at 31 °C. These results suggest that the antisense construct in the transgenic plants may be targeting the expression of the wild-type *RSW*1 gene at 21 °C.

Figure 7 shows the reduced inflorescence (bolt) height in antisense 35S-*RSW*1 plants compared to wild-type *A. thaliana* Columbia plants grown under identical conditions.

4. Cell wall carbohydrate analysis of antisense plants.

T3 plants which are homozygous for the 35S-RSW1 antisense construct are generated and the

content of cellulose therein is determined as described in Example 1. Plants expressing the antisense construct are shown to have significantly less cellulose in their cell walls, compared to wild-type plants. Additionally, the levels of non-crystalline β-1,4-glucan and starch are elevated in the cells of antisense plants, compared to otherwise isogenic plant lines which have not been transformed with the antisense genetic construct.

5. Antisense 35S-RSW1 mRNA expression levels in transgenic plants

Total RNA was extracted from 0.2g of leaf tissue derived from 33 kanamycin-resistant T1 plants containing the antisense 35S-*RSW1* genetic construct, essentially according to Longemann *et al.* (1986). Total RNA (25 μg) was separated on a 2.2M formaldehyde/agarose gel, blotted onto nylon filters and hybridized to a riboprobe comprising the sense strand sequence of the cDNA clone T20782. To produce the riboprobe, T7 RNA polymerase was used to transcribe sense RNA from a linearised plasmid template containing T20782, in the presence of [α-^{32}P]UTP. Hybridizations and subsequent washes were performed as described by Dolferus *et al.* (1994). Hybridized membranes were exposed to Phosphor screens (Molecular Dynamics, USA).

The levels of expression of the RSW1 antisense transcript were determined and compared to the level of fertility observed for the plant lines. As shown in Table 9, the level of antisense gene expression is correlated with the reduced fertility phenotype of the antisense plants. In 13 lines, a very high or high level of expression of the 35S-*RSW1* antisense gene was observed and, in 11 of these lines fertility was reduced. Only lines 2W and 3E which expressed high to very high levels of antisense mRNA, appeared to be fully fertile. In 12 lines which expressed medium levels of antisense mRNA, approximately one-half were fertile and one-half appeared to exhibit reduced fertility. In contrast, in 8 plant lines in which only a low or very low level of expression of the antisense 35S-RSW1 genetic construct was observed, a wild-type (i.e. fertile) phenotype was observed for all but one transgenic line, line 2R.

Data presented in Table 9 and Figure 7 indicate that the phenotype of the cellulose-deficient mutant *rsw1* may be reproduced by expressing antisense *RSW1* genetic constructs in transgenic

plants.

To confirm reduced cellulose synthesis and/or deposition in transgenic plants expressing the antisense *RSW1* gene, the level of cellulose is measured by the ^{14}C incorporation assay or as acetic/nitric acid insoluble material as described in Example 1 and compared to cellulose production in otherwise isogenic wild-type plants. Cellulose production in the transgenic plants is shown to be significantly reduced compared to wild-type plants. The severity of phenotype of the transgenic plants thus produced varies considerably, depending to some extent upon the level of inhibition of cellulose biosynthesis.

TABLE 9

LEVELS OF ANTISENSE GENE EXPRESSION AND FERTILITY IN T1 LINES OF ANTISENSE 35S-*RSW1* PLANTS

T1 PLANT LINE	ANTISENSE 35S-*RSW1* EXPRESSION	FERTILITY	T1 PLANT LINE	ANTISENSE 35S-*RSW1* EXPRESSION	FERTILITY
B	very high	sterile*	2H	medium	fertile
2B	very high	sterile*	C	medium	sterile*
3E	very high	fertile	F	medium	sterile*
2E	high	sterile*	2Q	medium	fertile
2K	high	sterile*	3P	medium	sterile*
2M	high	sterile*	3T	medium	fertile
2O	high	sterile*	5D	medium	sterile*
2P	high	sterile*	6A	medium	fertile
2W	high	fertile	8E	low	fertile
2Z	high	sterile*	2R	low	sterile*
3G	high	sterile*	7A	low	fertile
3Q	high	sterile*	7S	low	fertile
7Q	high	sterile*	7O	low	fertile
7N	medium	sterile*	7R	low	fertile
7G	medium	fertile	1B	very low	fertile
1C	medium	sterile*	2U	very low	fertile
2X	medium	sterile*			

*sterile phenotype not indicative of complete sterility, but that hand pollination at least, is required to obtain seed from such plants.

EXAMPLE 10

RSW1 RELATED SEQUENCES IN RICE PLANTS

To identify *RSW1* related nucleotide sequences in rice, a genetic sequence database was searched for nucleotide sequences which were closely-related to one or more of the *Arabidopsis thaliana RSW1* nucleotide sequences described in the preceding Examples. Rice EST S0542 (MAFF DNA bank, Japan) was identified, for which only a partial nucleotide sequences was available. Additionally, before the instant invention, there was no probable function attached to the rice EST S0542 sequence.

The present inventors have obtained the complete nucleotide sequence of clone S0542 and derived the amino acid sequence encoded therefor. The S0542 cDNA is only 1741bp in length and appears to be a partial cDNA clone because, although it comprises 100bp of 5'-untranslated sequence and contains the ATG start codon, it is truncated at 3'-end and, as a consequence encodes only the first 547 amino acid residues of the rice RSW1 or RSW1-like polypeptide. Based upon the length of the corresponding *Arabidopsis thaliana* RSW1 polypeptide (1081 amino acids), the rice RSW1 sequence set forth in SEQ ID NO:14 appears to contain approximately one-half of the complete amino acid sequence.

The N-terminal half of the rice RSW1 amino acid sequence is approximately 70% identical to the *Arabidopsis thaliana* RSW1 polypeptide set forth in SEQ ID NO:6, with higher homology (approximately 90%) occurring between amino acid residues 271-547 of the rice sequence. These data strongly suggest that S0542 is the rice homologue of the *A. thaliana RSW1* gene. Alignments of rice, *A. thaliana* and cotton RSW1 amino acid sequences are presented in Figures 9 and 10.

To isolate full-length cDNA clones and genomic clone equivalents of S0542 (this study and MAFF DNA bank, Japan) or D48636 (Pear *et al.*, 1996), cDNA and genomic clone libraries are produced using rice mRNA and genomic DNA respectively, and screened by hybridisation using the S0542 or D48636 cDNAs as a probe, essentially as described herein. Positive-

hybridising plaques are identified and plaque-purified, during further rounds of screening by hybridisation, to single plaques.

The rice clones are sequenced as described in the preceding Examples to determine the complete nucleotide sequences of the rice *RSW*1 genes and derived amino acid sequences therefor. Those skilled in the art will be aware that such gene sequences are useful for the production of transgenic plants, in particular transgenic cereal plants having altered cellulose content and/or quality, using standard techniques. The present invention extends to all such genetic sequences and applications therefor.

EXAMPLE 11

*RSW*1 RELATED SEQUENCES IN COTTON PLANTS

A ^{32}P-labelled *RSW*1 PCR fragment was used to screen approximately 200,000 cDNA clones in a cotton fibre cDNA library. The *RSW*1 PCR probe was initially amplified from *Arabidopsis thaliana* wild type cDNA using the primers 2280-F and csp1-R described in the preceding Examples, and then re-amplified using the primer combination 2370-F/csp1-R, also described in the preceding Examples.

Hybridisations were carried out under low stringency conditions at 55°C.

Six putative positive-hybridising plaques were identified in the first screening round. Using two further rounds of screening by hybridisation, four of these plaques were purified to single plaques. Three plaques hybridise very strongly to the *RSW*1 probe while the fourth plaque hybridises less intensely.

We conclude that the positive-hybridising plaques which have been purified are strong candidates for comprising cotton *RSW*1 gene sequences or *RSW*1-like gene sequences. Furthermore, the cotton cDNAs may encode the catalytic subunit of cellulose synthase,

because the subunit protein architecture of cellulose synthase appears to be highly conserved among plants as highlighted in the preceding Example.

Furthermore, a Southern blot of cotton genomic DNA digested with *Bgl*II was hybridised with the 5' end of the *RSW*1 cDNA, under low stringency hybridisation conditions at 55°C. Results are presented in Figure 11. These data demonstrate that RSW1-related sequences exist in the cotton genome.

The cotton cDNA clones described herein are sequenced as described in the preceding Examples and used to produce transgenic cotton plants having altered fibre characteristics. The cDNAs are also used to genetically alter the cellulose content and/or quality of other plants, using standard techniques.

EXAMPLE 12

*RSW*1 RELATED SEQUENCES IN *EUCALYPTUS SSP.*

Putative *Eucalyptus ssp.* cellulose synthase catalytic subunit gene fragments were obtained by amplification using PCR. DNA primers were designed to conserved amino acid residues found in the *Arabidopsis thaliana RSW*1 and 12C4 amino acid sequences. Three primers were used for PCR. The primers are listed below:

pcsF-I 5'- A A/G A A G A T I G A C/T T A C/T C/T T I A A A/G G A C/T A A-3'
pcsR-II 5'-A T I G T I G G I G T I C G/T A/G T T C/T T G A/T/G/C C T/G A/T/C/G C C -3'
pcsF-II 5'- G C I A T G A A A A/G A/C G I G A I T A C/T G A A/G G A -3'

Using standard PCR conditions (50°C annealing temperature) and solutions, the primer sets pcsF-I/pcsR-II and pcsF-II/pcsR-II were used to amplify genetic sequences from pooled *Eucalyptus ssp.* cDNA. In the first reaction primers pcsF-I and pcsR-II were used to generate a fragment approximately 700 bp in length. In the second PCR reaction, which used primers pcsF-II and pcsR-II, a fragment estimated to 700 bp was obtained. The sizes of the PCR

fragments are within the size range estimated for the corresponding *Arabidopsis thaliana* sequences.

We conclude that the amplified *Eucalyptus ssp.* PCR fragments are likely to be related to the *Arabidopsis thaliana RSW1* gene and may encode at least a part of the *Eucalyptus ssp.* cellulose synthase catalytic subunit.

The *Eucalyptus ssp.* PCR clones described herein are sequenced as described in the preceding Examples and used to isolate the corresponding full-length *Eucalyptus ssp* cDNAs and genomic gene equivalents. Those skilled in the art will be aware that such gene sequences are useful for the production of transgenic plants, in particular transgenic *Eucalyptus ssp* plants having altered cellulose content and/or quality, using standard techniques. The present invention extends to all such genetic sequences and applications therefor.

EXAMPLE 13

NON-CRYSTALLINE B-1,4-GLUCAN AS A MODIFIER OF CELL WALL PROPERTIES

The properties of plant cell walls depend on the carbohydrates, proteins and other polymers of which they are composed and the complex ways in which they interact. Increasing the quantities of non-crystalline β-1,4-glucan in cell walls affects those wall properties which influence mechanical, nutritional and many other qualities as well as having secondary consequences resulting from the diversion of carbon into non-crystalline glucan at the expense of other uses. To illustrate one of these effects, we investigated the ability of the non-crystalline glucan to hydrogen bond to other wall components particularly cellulose in the way that has been shown to be important for wall mechanics.

Hemicelluloses such as xyloglucans cross-link cellulose microfibrils by hydrogen bonding to the microfibril surface (Levy *et al.*, 1991). Since the β-1,4-glucan backbone of xyloglucan is

thought to be responsible for hydrogen bonding (with the xylose, galactose and fucose substitutions limiting the capacity to form further hydrogen bonds) we can expect the non-crystalline β-1,4-glucan also to have a capacity to hydrogen bond and cross link cellulose. The effectiveness of strong alkalis in extracting xyloglucans is thought to relate to their disruption of the hydrogen bonds with cellulose (Hayashi and MacLachlan, 1984).

To demonstrate that the non-crystalline β-1,4-glucan forms similar associations with the cellulose microfibrils, we examined whether the 4 M KOH fraction, extracted from shoots of the *rsw*1 mutant and from wild type *RSW*1 plants, contained non-crystalline glucan in addition to xyloglucan. The non-crystalline glucan was separated from xyloglucan in the 4 M KOH extract by dialysing the neutralised extract against distilled water and centrifuging at 14000 g for 1 hour. The pellet was shown to be a pure β-1,4-glucan by using the methods for monosaccharide analysis, methylation analysis and enzyme digestion used to characterise the glucan in the ammonium oxalate fraction (see Example 1).

Table 10 shows the presence of substantial quantities of glucan recovered in pure form in the pellet from 4 M KOH fractions extracted from the overproducing *rsw*1 mutant of *Arabidopsis thaliana*. These data also demonstrate the presence of smaller quantities of non-crystalline β-1,4-glucan in the 4 M KOH fraction from wild type plants, compared to *rsw*1, particularly when grown at 31 °C.

TABLE 10

Glucose contents* of 4M KOH fractions from shoots of wild-type and *rsw*1 mutant *Arabidopsis thaliana* plants

Glucose fraction	wild-type		*rsw*1 mutant	
	21°C	31°C	21°C	31°C
xyloglucan and non-crystalline glucan in whole extract	36.4	56.9	27.1	93.1
non-crystalline glucan in pellet	7.8	20.5	7.6	56.0

*, nmol glucose/ mg plant dry weight after TFA hydrolysis

The monosaccharide composition of the supernatant remaining after centrifugation was determined after TFA hydrolysis. These data, and data from methylation analysis, are consistent with the supernatant being a relatively pure xyloglucan. The supernatant was free of glucan, because no glucose could be released by the endocellulase/β-glucosidase mixture that released glucose from β-1,4-glucan.

The presence of both non-crystalline β-1,4-glucan and xyloglucan in the 4 M KOH fraction, when taken together with the implications from structural predictions (Levy *et al*, 1991), is consistent with some of the non-crystalline β-1,4-glucan in the wall hydrogen bonding to cellulose microfibrils in similar fashion to the β-1,4-glucan backbone of xyloglucan.

The cross linking provided when xyloglucans and other hemicelluloses bind to two or more microfibrils is an important determinant of the mechanical properties of cellulosic walls (Hayashi, 1989). The effects of increasing the amounts of non-crystalline β-1,4-glucan in walls are likely to be greatest in walls which otherwise possess relatively low levels of cross linking as a result of high ratios of cellulose: hemicelluloses. Such conditions are common in secondary walls including those of various fibres, and the cellulose:hemicellulose ratio is particularly high in cotton fibres.

The effects on wall mechanical properties of overproducing non-crystalline glucan are shown by transforming plants with the mutant allele of *rsw1* (SEQ ID NO:11) operably under the control of either the *RSW1* promoter derived from SEQ ID NO:3 or SEQ ID NO:4 or alternatively, an appropriate constitutive promoter such as the CaMV 35S promoter. Production of non-crystalline glucan is quantified by fractionating the cell walls using the methods described above to show in particular that non-crystalline glucan is recovered in the 4 M KOH fraction. Mechanical properties of the cell walls are measured using standard methods for fibre analysis to study parameters such as stress-strain curves, and breaking strain, amongst other properties.

EXAMPLE 14
OVER-EXPRESSION OF CELLULOSE SYNTHASE IN TRANSGENIC PLANTS

Three strategies are employed to over-express cellulose synthase in *Arabidopsis thaliana* plants.

In the first strategy, the CaMV 35S promoter sequence is operably connected to the full-length cellulose synthase cDNA which is obtainable by primer extension of SEQ ID NO:1. This is achievable by cloning the full-length cDNA encoding cellulose synthase, in the sense orientation, between the CaMV 35S promoter or other suitable promoter operable in plants and the nopaline synthase terminator sequences of the binary plasmid pBI121.

In the second strategy, the coding part of the genomic gene is cloned, in the sense orientation, between the CaMV 35S promoter and the nopaline synthase terminator sequences of the binary plasmid pBI121.

In the third strategy, the 23H12 binary cosmid clone or the derivative pRSW1, containing the cellulose synthase gene sequence operably under the control of the cellulose synthase gene promoter and terminator sequences is prepared in a form suitable for transformation of plant tissue.

For *Agrobacterium*-mediated tissue transformation, binary plasmid constructs discussed *supra* are transformed into *Agrobacterium tumefaciens* strain AGL1 or other suitable strain. The recombinant DNA constructs are then introduced into wild type *Arabidopsis thaliana* plants (Columbia ecotype), as described in the preceding Examples.

Alternatively, plant tissue is directly transformed using the vacuum infiltration method described by Beshtold *et al.* (1993).

The transgenic plants thus produced exhibit a range of phenotypes, partly because of position effects and variable levels of expression of the cellulose synthase transgene.

Cellulose content in the transgenic plants and isogenic untransformed control plants is determined by the ^{14}C incorporation assay or as acetic/nitric acid insoluble material as described in Example 1. In general, the level of cellulose deposition and rates of cellulose biosynthesis in the transgenic plants are significantly greater than for untransformed control plants.

Furthermore, in some cases, co-supression leads to mimicry of the *rsw*1 mutant phenotype.

EXAMPLE 15

SITE-DIRECTED MUTAGENESIS OF THE *RSW*1 GENE

The nucleotide sequence of the *RSW*1 gene contained in 23H12 is mutated using site-directed mutagenesis, at several positions to alter its catalytic activity or substrate affinity or glucan properties. In one example, the RSW1 gene is mutated to comprise one or more mutations present in the mutant *rsw*1 allele.

The mutated genetic sequences are cloned into binary plasmid described in the preceding Examples, in place of the wild-type sequences. Plant tissue obtained from both wild-type *Arabidopsis thaliana* (Columbia) plants and *A. thaliana rsw*1 plants is transformed as described herein and whole plants are regenerated.

Control transformations are performed using the wild-type cellulose synthase gene sequence.

EXAMPLE 16

PHENOTYPES OF PLANTS EXPRESSING MUTATED *RSW1* GENES

Plants transformed with genetic constructs described in Example 15 (and elsewhere) are categorised initially on the basis of number of transgene copies, to eliminate variability arising therefrom. Plants expressing single copies of different transgenes are analysed further for cell wall components, including cellulose, non-crystalline β-1,4-glucan polymer, starch and carbohydrate content.

1. Cellulose content

Cellulose content in the transgenic plants is determined by the ^{14}C incorporation assay as described in Example 1. Cell walls are prepared, fractionated and the monosaccharide composition of individual fractions determined as in Example 1.

2. Non-crystalline β-1,4-glucan content

Transgenic plants expressing the *rsw1* mutant allele exhibit a higher level of non-crystalline, and therefore extractable, β-1,4-glucan in cell walls compared to plants expressing an additional copy of the wild-type *RSW1* allele. Thus, it is possible to change the crystallinity of the β-1,4-glucan chains present in the cell wall by mutation of the wild-type *RSW1* allele.

3. Starch content

Transgenic plants are also analysed to determine the effect of mutagenesis of the *RSW1* gene on the level of starch deposited in their roots. The quantity of starch present in material prepared from the crude wall fraction is determined using the anthrone/H_2SO_4 method described in Example 1. The data show that mutating the *RSW1* gene to the mutant *rsw1* allele increases starch deposition. This demonstrates that the gene can be used to alter the partitioning of carbon into carbohydrates other than cellulose.

4.Cell wall composition

The cell wall composition of transgenic plant material is also analysed. Wild type and *rsw1*

and transgenic seedlings are grown for 2 d at 21°C and then kept for a further 5 d at either 21°C or 31°C. With transfer to 31°C when the seed has scarcely germinated, the wall composition at final harvest largely reflects the operation of the mutated *rsw*1 gene product at its restrictive temperature. Cell wall fractionation is carried out in similar fashion to that described for the ^{14}C-experiment (Example 1) and the monosaccharide composition of each fraction is quantified by GC/MS after hydrolysis with trifluoroacetic acid or, in the case of crystalline cellulose, H_2SO_4.

In some transgenic plants in which the *RSW*1 gene is mutated, the monosaccharide composition is comparable to that observed for homozygous *rsw*1 plants, at least in some cases, confirming that there is a major reduction in the quantity of crystalline cellulose in the final, acid insoluble fraction. Thus, mutation of the *RSW*1 gene can be performed to produce changes in the composition of plant cell walls.

EXAMPLE 17

CHEMICAL MODIFICATION OF THE *RSW*1 GENE TO MANIPULATE CELLULOSE PRODUCTION AND PLANT CELL WALL CONTENT.

As demonstrated in the preceding Examples, the *RSW*1 gene is involved in cellulose production and the manipulation of cell wall content.

In the present Example, to identify novel phenotypes and gene sequences important for the normal functioning of the cellulose synthase gene, the *RSW*1 gene is modified *in planta*, using the chemical mutagen EMS. The mutant plants are identified following germination and the modified *RSW*1 genes are isolated and characterised at the nucleotide sequence level. A sequence comparison between the mutant gene sequences and the wild type sequence reveals nucleotides which encode amino acids important to the normal catalytic activity of the cellulose synthase enzyme, at least in *Arabidopsis thaliana* plants.

This approach thus generates further gene sequences of utility in the modification of cellulose

content and properties in plants.

EXAMPLE 18

DISCUSSION

Five pieces of evidence make a compelling case that the RSW1 gene product encodes the catalytic subunit of cellulose synthase:

1. The *rsw1* mutation selectively inhibits cellulose synthesis and promotes accumulation of a non-crystalline β-1,4-glucan;
2. The *rsw1* mutation removes cellulose synthase complexes from the plasma membrane, providing a plausible mechanism for reduced cellulose accumulation and placing the RSW1 product either in the complexes or interacting with them;
3. The D,D,D,QXXRW signature identifies the RSW1 gene product as a processive glycosyl transferase enzyme (Saxena, 1995);
4. The wild type allele corrects the temperature sensitive phenotype of the *rsw1* mutant; and
5. Antisense expression of the *RSW1* in transgenic plants grown at 21 °C reproduces some of the phenotype of *rsw1* which is observed following growth at 31 °C.

Consistent with the plasma membrane location expected for a catalytic subunit, the putative 122 kDa RSW1 product contains 8 predicted membrane-spanning regions. Six of these regions cluster near the C-terminus (Figure 10), separated from the other two by a domain that is probably cytoplasmic and has the weak sequence similarities to prokaryotic glycosyl transferases (Wong, 1990; Saxena, 1990 ; Matthyse, 1995; Sofia, 1994 ; Kutish, 1996).

RSW1 therefore qualifies as a member of the large family of *Arabidopsis thaliana* genes whose members show weak similarities to bacterial cellulose synthase. RSW1 is the first member of that family to be rigorously identified as an authentic cellulose synthase. Among the diverse genes in *A. thaliana*, at least two genes show very strong sequence similarities to the *RSW1* gene and are most likely members of a highly conserved sub-family involved in

cellulose synthesis. The closely related sequences come from cosmid 12C4, a partial genomic clone cross-hybridising with EST T20782 designated *Ath*-A, and from a full length cDNA designated *Ath*-B.

Ath-A resembles *RSW*1 (SEQ ID NO:5) at its N-terminus whereas *Ath*-B starts 22 amino acid residues downstream [Figure 8 and Figure 9(i), (ii) and (iii)]. Closely related sequences in other angiosperms are the rice EST S0542 [Figure 9(i), (ii) and (iii)], which resembles the polypeptides encoded by *RSW*1 and *Ath*-A and the cotton *cel*A1 gene (Pear, 1996) at the N-terminus.

The *Arabidopsis thaliana*, rice and cotton genes have regions of very high sequence similarity interspersed with variable regions (Figures 9 and 10). Most of the highest conservation among those gene products occurs in their central cytoplasmic domain where the weak similarities to the bacterial cellulose synthase occur. The N-terminal region that precedes the first membrane spanning region is probably also cytoplasmic but shows many amino acid substitutions as well as sequences in RSW1 that have no counterpart in some of the other genes as already noted for *cel*A. An exception to this is a region comprising 7 cysteine residues with highly conserved spacings (Figure 10). This is reminiscent of regions suggested to mediate protein-protein and protein-lipid interactions in diverse proteins including transcriptional regulators and may account for the striking sequence similarity between this region of RSW1 and two putative soybean bZIP transcription factors (Genbank SOYSTF1A and 1B).

In conclusion, the chemical and ultrastructural changes seen in the cellulose-deficient mutant combine with gene cloning and complementation of the mutant to provide strong evidence that the *RSW*1 locus encodes the catalytic subunit of cellulose synthase. Accumulation of non-crystalline β- 1,4-glucan in the shoot of the *rsw*1 mutant suggests that properties affected by the mutation are required for glucan chains to assemble into microfibrils. Whilst not being bound by any theory or mode of action, a key property may be the aggregation of catalytic subunits into plasma membrane rosettes. At the restrictive temperature, mutant synthase

complexes disassemble to monomers (or smaller oligomers) that are undetectable by freeze etching. At least in the shoot, the monomers seem to remain biosynthetically active but their β-1,4-glucan products fail to crystallise into microfibrils probably because the chains are growing from dispersed sites. Crystallisation into microfibrils, with all its consequences for wall mechanics and morphogenesis, therefore may depend upon catalytic subunits remaining aggregated as plasma membrane rosettes.

Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations or any two or more of said steps or features.

REFERENCES

1. **An *et al.*** (1985) EMBO J. 4:277-284.
2. **Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K.** (1987) Current Protocols in Molecular Biology, Wiley Interscience (ISBN 047140338).
3. **Baskin *et al.*** (1992) Aust. J. Plant Physiol. **19**:427-437.
4. **Bechtold *et al.*** (1993) Planta **316**:1194-1199.
5. **Bevan, M.** (1984) Nucl. Acids Res. **12**, 8711-8721.
6. **Chang, C. *et al*** (1988) Proc. Natl. Acad. Sci. (USA) **85**, 6856-6860.
7. **Cole*et al.*** (1985) In Monoclonal antibodies in cancer therapy, Alan R. Bliss Inc., pp 77-96;
8. **Crossway *et al.*** (1986) Mol. Gen. Genet. **202**:179-185.

Datla, R.S.S., Hammerlindl, J.K., Panchuk, B., Pelcher, L.E. and Keller, W. '992) Gene, **211**, 383-384.

et al. (1980) Proc. Natl. Acad. Sci. (USA) **77**, 7347-7351.

'hersheim, P. and Darvill, A.G. (1991) Carb. Res. **210**, 311-317.

12. **Dolferus, R., Jacobs, M., Peacock, W.J. and Dennis, E.S.** (1994) Plant Physiol. **105**, 1075-1087.

13. **Fromm *et al.*** (1985) Proc. Natl. Acad. Sci. (USA) **82**:5824-5828.

14. **Haseloff, J., and Gerlach, W.L.** (1988) Nature **334**:586-594

15. **Hayashi** (1989) Ann Rev Plant Physiol. Plant Molecular Biol. **40**, 139-168.

16. **Hayashi and MacLachlan** (1984) Plant Physiol. **75**, 596-604.

17. **Herrera-Estrella *et al.*** (1983a) Nature **303**:209-213.

18. **Herrera-Estrella *et al.*** (1983b) EMBO J. **2**:987-995

19. **Herrera-Estrella *et al.*** (1985) In: Plant Genetic Engineering, Cambridge University Press. NY, pp 63-93.

20. **Herth, W.** (1985) Planta **164**, 12-21.

21. **Huse *et al.*** (1989) Science **246**: 1275-1281.

22. **Kohler and Milstein** (1975) Nature, **256**: 495-499.

23. **Konieczny, A. and Ausubel, F.** (1993) Plant J. **4**, 403-410.

24. **Kozbor *et al.*** (1983) Immunol. Today **4**: 72.

25. **Lazo, G.R., Stein, P.A. and Ludwig, R.A.** (1991). Bio/technology 9,963-967.

26. **Levy *et al.*** (1991) Plant Journal **1**, 195-215.

27. **Longemann, J, Schell, J. and Willmitzer, L.** (1987). Anal. Biochem. **163**, 16-20.

28. **Matthyse, A.G., White, S., and Lightfoot, R.** (1995) J. Bacteriol. **177**, 1069-1075.

29. **McPherson *et al.*** (1991) In: PCR:A Practical Approach. IRL Press. Oxford.

30. **Nam, H.-G., *et al.*** (1989) Plant Cell **1**, 699-705.

31. **Needs, P.W. and Selvendran, R.R.** (1993) Phytochem. Anal. **4**, 210-216.

32. **Paszkowski *et al.*** (1984) EMBO J. **3**:2717-2722.

33. **Pear, J.R.,*et al*** (1996) Proc. Natl. Acad. Sci. (USA) **93**, 12637-12642.

34. **Saxena *et al.*** (1990) Plant Mol. Biol. **15**, 673-683.

35. **Saxena *et al.*** (1995) J.Bacteriol. **177**: 1419-1424.

36. **Sofia, H.J., *et al*** (1994) Nucl. Acids Res. **22**, 2576-2586.

37. **Southern, E.M.** (1975). J. Mol. Biol. **98**, 503-517.

38. **Updegraph, D.J.**. (1969) Analyt. Bioch. **32**: 429-424.

39. **Wong, H.C. *et al*** (1990) Proc. Natl. Acad. Sci. (USA) **87**:8130-8134.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Australian National University and the Commonwealth Scientific and
Industrial Research Organisation

(ii) TITLE OF INVENTION: Manipulation of plant cellulose

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Davies Collison Cave Patent Attorneys
(B) STREET: 1, Little Collins Street
(C) CITY: Melbourne
(D) STATE: Victoria
(E) COUNTRY: Australia
(F) ZIP: 3000

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT INTERNATIONAL
(B) FILING DATE:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: AU PO0699
(B) FILING DATE: 27-JUN-1996

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: SLATTERY, JOHN M

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 61-3-9254-2777
(B) TELEFAX: 61-3-9254-2770

(C) TELEX: AA31787

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2248 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana

(vii) IMMEDIATE SOURCE:
(B) CLONE: EST T20782

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1887

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGA GCT ATG AAG AGA GAG TAT GAA GAG TTT AAA GTG AGG ATA AAT GCT        48
Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala
 1               5                  10                  15

CTT GTT GCC AAA GCA CAG AAA ATC CCT GGA GAA GGC TGG ACA ATG CAG        96
Leu Val Ala Lys Ala Gln Lys Ile Pro Gly Glu Gly Trp Thr Met Gln
            20                  25                  30

GAT GGT ACT CCC TGG CCT GGT AAC AAC ACT AGA GAT CAT CCT GGA ATG       144
Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp His Pro Gly Met
        35                  40                  45
```

```
ATA CAG GTG TTC TTA GGC CAT AGT GGG GGT CTG GAT ACC GAT GGA AAT      192
Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp Thr Asp Gly Asn
     50                  55                  60

GAG CTG CCT AGA CTC ATC TAT GTT TCT CGT GAA AAG CGG CCT GGA TTT      240
Glu Leu Pro Arg Leu Ile Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe
 65                  70                  75                  80

CAA CAC CAC AAA AAG GCT GGA GCT ATG AAT GCA TCG ATC CGT GTA TCT      288
Gln His His Lys Lys Ala Gly Ala Met Asn Ala Ser Ile Arg Val Ser
                 85                  90                  95

GCT GTT CTT ACC AAT GGA GCA TAT CTT TTG AAC GTG GAT TGT GAT CAT      336
Ala Val Leu Thr Asn Gly Ala Tyr Leu Leu Asn Val Asp Cys Asp His
            100                 105                 110

TAC TTT AAT AAC AGT AAG GCT ATT AAA GAA GCT ATG TGT TTC ATG ATG      384
Tyr Phe Asn Asn Ser Lys Ala Ile Lys Glu Ala Met Cys Phe Met Met
        115                 120                 125

GAC CCG GCT ATT GGA AAG AAG TGC TGC TAT GTC CAG TTC CCT CAA CGT      432
Asp Pro Ala Ile Gly Lys Lys Cys Cys Tyr Val Gln Phe Pro Gln Arg
    130                 135                 140

TTT GAC GGT ATT GAT TTG CAC GAT CGA TAT GCC AAC AGG AAT ATA GTC      480
Phe Asp Gly Ile Asp Leu His Asp Arg Tyr Ala Asn Arg Asn Ile Val
145                 150                 155                 160

TTT TTC GAT ATT AAC ATG AAG GGG TTG GAT GGT ATC CAC GGT CCA GTA      528
Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Ile His Gly Pro Val
                165                 170                 175

TAT GTG GGT ACT GGT TGT TGT TTT AAT AGG CAG GCT CTA TAT GGG TAT      576
Tyr Val Gly Thr Gly Cys Cys Phe Asn Arg Gln Ala Leu Tyr Gly Tyr
            180                 185                 190

GAT CCT GTT TTG ACG GAA GAA GAT TTA GAA CCA AAT ATT ATT GTC AAG      624
Asp Pro Val Leu Thr Glu Glu Asp Leu Glu Pro Asn Ile Ile Val Lys
        195                 200                 205
```

```
AGC TGT TGC GGG TCA AGG AAG AAA GGT AAA AGT AGC AAG AAG TAT AAC      672
Ser Cys Cys Gly Ser Arg Lys Lys Gly Lys Ser Ser Lys Lys Tyr Asn
    210                 215                 220

TAC GAA AAG AGG AGA GGC ATC AAC AGA AGT GAC TCC AAT GCT CCA CTT      720
Tyr Glu Lys Arg Arg Gly Ile Asn Arg Ser Asp Ser Asn Ala Pro Leu
225                 230                 235                 240

TTC AAT ATG GAG GAC ATC GAT GAG GGT TTT GAA GGT TAT GAT GAT GAG      768
Phe Asn Met Glu Asp Ile Asp Glu Gly Phe Glu Gly Tyr Asp Asp Glu
                245                 250                 255

AGG TCT ATT CTA ATG TCC CAG AGG AGT GTA GAG AAG CGT TTT GGT CAG      816
Arg Ser Ile Leu Met Ser Gln Arg Ser Val Glu Lys Arg Phe Gly Gln
            260                 265                 270

TCG CCG GTA TTT ATT GCG GCA ACC TTC ATG GAA CAA GGC GGC ATT CCA      864
Ser Pro Val Phe Ile Ala Ala Thr Phe Met Glu Gln Gly Gly Ile Pro
        275                 280                 285

CCA ACA ACC AAT CCC GCT ACT CTT CTG AAG GAG GCT ATT CAT GTT ATA      912
Pro Thr Thr Asn Pro Ala Thr Leu Leu Lys Glu Ala Ile His Val Ile
    290                 295                 300

AGC TGT GGT TAC GAA GAC AAG ACT GAA TGG GGC AAA GAG ATT GGT TGG      960
Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp Gly Lys Glu Ile Gly Trp
305                 310                 315                 320

ATC TAT GGT TCC GTG ACG GAA GAT ATT CTT ACT GGG TTC AAG ATG CAT     1008
Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly Phe Lys Met His
                325                 330                 335

GCC CGG GGT TGG ATA TCG ATC TAC TGC AAT CCT CCA CGC CCT GCG TTC     1056
Ala Arg Gly Trp Ile Ser Ile Tyr Cys Asn Pro Pro Arg Pro Ala Phe
            340                 345                 350

AAG GGA TCT GCA CCA ATC AAT CTT TCT GAT CGT TTG AAC CAA GTT CTT     1104
Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu Asn Gln Val Leu
        355                 360                 365
```

```
CGA TGG GCT TTG GGA TCT ATC GAG ATT CTT CTT AGC AGA CAT TGT CCT      1152
Arg Trp Ala Leu Gly Ser Ile Glu Ile Leu Leu Ser Arg His Cys Pro
    370                 375                 380

ATC TGG TAT GGT TAC CAT GGA AGG TTG AGA CTT TTG GAG AGG ATC GCT      1200
Ile Trp Tyr Gly Tyr His Gly Arg Leu Arg Leu Leu Glu Arg Ile Ala
385                 390                 395                 400

TAT ATC AAC ACC ATC GTC TAT CCT ATT ACA TCC ATC CCT CTT ATT GCG      1248
Tyr Ile Asn Thr Ile Val Tyr Pro Ile Thr Ser Ile Pro Leu Ile Ala
                405                 410                 415

TAT TGT ATT CTT CCC GCT TTT TGT CTC ATC ACC GAC AGA TTC ATC ATA      1296
Tyr Cys Ile Leu Pro Ala Phe Cys Leu Ile Thr Asp Arg Phe Ile Ile
            420                 425                 430

CCC GAG ATA AGC AAC TAC GCG AGT ATT TGG TTC ATT CTA CTC TTC ATC      1344
Pro Glu Ile Ser Asn Tyr Ala Ser Ile Trp Phe Ile Leu Leu Phe Ile
        435                 440                 445

TCA ATT GCT GTG ACT GGA ATC CTG AAA CTG AAA TGG AAC GGT GTG AGC      1392
Ser Ile Ala Val Thr Gly Ile Leu Lys Leu Lys Trp Asn Gly Val Ser
    450                 455                 460

ATT GAG GAT TGG TGG AGG AAC AAC CAG TTC TGG GTC ATT GGT GGC ACA      1440
Ile Glu Asp Trp Trp Arg Asn Asn Gln Phe Trp Val Ile Gly Gly Thr
465                 470                 475                 480

TCC ACC CAT CTT TTT GCT GTC TTC CAA GGT CTA CTT AAG GTT CTT GCT      1488
Ser Thr His Leu Phe Ala Val Phe Gln Gly Leu Leu Lys Val Leu Ala
                485                 490                 495

GGT ATC AAC ACC AAC TTC ACC GTT ACA TCT AAA GCC ACA AAC AAA AAT      1536
Gly Ile Asn Thr Asn Phe Thr Val Thr Ser Lys Ala Thr Asn Lys Asn
            500                 505                 510

GGG GAT TTT GCA AAA CTC TAC ATC TTC AAA TGG ACA GCT CTT CTC ATT      1584
Gly Asp Phe Ala Lys Leu Tyr Ile Phe Lys Trp Thr Ala Leu Leu Ile
        515                 520                 525
```

```
CCA CCA ACC ACC GTC CTA CTT GTG AAC CTC ATA GGC ATT GTG GCT GGT      1632
Pro Pro Thr Thr Val Leu Leu Val Asn Leu Ile Gly Ile Val Ala Gly
    530                 535                 540

GTC TCT TAT GCT GTA AAC AGT GGC TAC CAG TCG TGG GGT CCG CTT TTC      1680
Val Ser Tyr Ala Val Asn Ser Gly Tyr Gln Ser Trp Gly Pro Leu Phe
545                 550                 555                 560

GGG AAG CTC TTC TTC GCC TTA TGG GTT ATT GCC CAT CTC TAC CCT TTC      1728
Gly Lys Leu Phe Phe Ala Leu Trp Val Ile Ala His Leu Tyr Pro Phe
                565                 570                 575

TTG AAA GGT CTG TTG GGA AGA CAA AAC CGA ACA CCA ACC ATC GTC ATT      1776
Leu Lys Gly Leu Leu Gly Arg Gln Asn Arg Thr Pro Thr Ile Val Ile
            580                 585                 590

GTC TGG TCT GTT CTT CTC GCC TCC ATC TTC TCG TTG CTT TGG GTC AGG      1824
Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu Leu Trp Val Arg
        595                 600                 605

ATC AAT CCC TTT GTG GAC GCC AAT CCC AAT GCC AAC AAC TTC AAT GGC      1872
Ile Asn Pro Phe Val Asp Ala Asn Pro Asn Ala Asn Asn Phe Asn Gly
    610                 615                 620

AAA GGA GGT GTC TTT TAGACCCTAT TTATATACTT GTGTGTGCAT ATATCAAAAA      1927
Lys Gly Gly Val Phe
625

CGCGCAATGG GAATTCCAAA TCATCTAAAC CCATCAAACC CCAGTGAACC GGGCAGTTAA    1987

GGTGATTCCA TGTCCAAGAT TAGCTTTCTC CGAGTAGCCA GAGAAGGTGA AATTGTTCGT    2047

AACACTATTG TAATGATTTT CCAGTGGGGA AGAAGATGTG GACCCAAATG ATACATAGTC    2107

TACAAAAAGA ATTAGTTATA ACTTTCTTAT ATTTATTTTA TTTAAAGCTT GTTAGACTCA    2167

CACTTATGTA ATGTTGGAAC TTGTTGTCCT AAAAAGGGAT TGGAGTTTTC TTTTTATCTA    2227

AGAATCTGAA GTTTATATGC T                                              2248
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 629 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala
 1               5                  10                  15

Leu Val Ala Lys Ala Gln Lys Ile Pro Gly Glu Gly Trp Thr Met Gln
            20                  25                  30

Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp His Pro Gly Met
        35                  40                  45

Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp Thr Asp Gly Asn
    50                  55                  60

Glu Leu Pro Arg Leu Ile Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe
65                  70                  75                  80

Gln His His Lys Lys Ala Gly Ala Met Asn Ala Ser Ile Arg Val Ser
                85                  90                  95

Ala Val Leu Thr Asn Gly Ala Tyr Leu Leu Asn Val Asp Cys Asp His
            100                 105                 110

Tyr Phe Asn Asn Ser Lys Ala Ile Lys Glu Ala Met Cys Phe Met Met
        115                 120                 125

Asp Pro Ala Ile Gly Lys Lys Cys Cys Tyr Val Gln Phe Pro Gln Arg
    130                 135                 140

Phe Asp Gly Ile Asp Leu His Asp Arg Tyr Ala Asn Arg Asn Ile Val
145                 150                 155                 160
```

Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Ile His Gly Pro Val
165 170 175

Tyr Val Gly Thr Gly Cys Cys Phe Asn Arg Gln Ala Leu Tyr Gly Tyr
180 185 190

Asp Pro Val Leu Thr Glu Glu Asp Leu Glu Pro Asn Ile Ile Val Lys
195 200 205

Ser Cys Cys Gly Ser Arg Lys Lys Gly Lys Ser Ser Lys Lys Tyr Asn
210 215 220

Tyr Glu Lys Arg Arg Gly Ile Asn Arg Ser Asp Ser Asn Ala Pro Leu
225 230 235 240

Phe Asn Met Glu Asp Ile Asp Glu Gly Phe Glu Gly Tyr Asp Asp Glu
245 250 255

Arg Ser Ile Leu Met Ser Gln Arg Ser Val Glu Lys Arg Phe Gly Gln
260 265 270

Ser Pro Val Phe Ile Ala Ala Thr Phe Met Glu Gln Gly Gly Ile Pro
275 280 285

Pro Thr Thr Asn Pro Ala Thr Leu Leu Lys Glu Ala Ile His Val Ile
290 295 300

Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp Gly Lys Glu Ile Gly Trp
305 310 315 320

Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly Phe Lys Met His
325 330 335

Ala Arg Gly Trp Ile Ser Ile Tyr Cys Asn Pro Pro Arg Pro Ala Phe
340 345 350

Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu Asn Gln Val Leu
355 360 365

Arg Trp Ala Leu Gly Ser Ile Glu Ile Leu Leu Ser Arg His Cys Pro
370 375 380

Ile Trp Tyr Gly Tyr His Gly Arg Leu Arg Leu Leu Glu Arg Ile Ala
385 390 395 400

Tyr Ile Asn Thr Ile Val Tyr Pro Ile Thr Ser Ile Pro Leu Ile Ala
405 410 415

Tyr Cys Ile Leu Pro Ala Phe Cys Leu Ile Thr Asp Arg Phe Ile Ile
420 425 430

Pro Glu Ile Ser Asn Tyr Ala Ser Ile Trp Phe Ile Leu Leu Phe Ile
435 440 445

Ser Ile Ala Val Thr Gly Ile Leu Lys Leu Lys Trp Asn Gly Val Ser
450 455 460

Ile Glu Asp Trp Trp Arg Asn Asn Gln Phe Trp Val Ile Gly Gly Thr
465 470 475 480

Ser Thr His Leu Phe Ala Val Phe Gln Gly Leu Leu Lys Val Leu Ala
485 490 495

Gly Ile Asn Thr Asn Phe Thr Val Thr Ser Lys Ala Thr Asn Lys Asn
500 505 510

Gly Asp Phe Ala Lys Leu Tyr Ile Phe Lys Trp Thr Ala Leu Leu Ile
515 520 525

Pro Pro Thr Thr Val Leu Leu Val Asn Leu Ile Gly Ile Val Ala Gly
530 535 540

Val Ser Tyr Ala Val Asn Ser Gly Tyr Gln Ser Trp Gly Pro Leu Phe
545 550 555 560

Gly Lys Leu Phe Phe Ala Leu Trp Val Ile Ala His Leu Tyr Pro Phe
565 570 575

Leu Lys Gly Leu Leu Gly Arg Gln Asn Arg Thr Pro Thr Ile Val Ile
580 585 590

Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu Leu Trp Val Arg
595 600 605

Ile Asn Pro Phe Val Asp Ala Asn Pro Asn Ala Asn Asn Phe Asn Gly
610 615 620

Lys Gly Gly Val Phe
625

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8411 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia (wild-type)

(vii) IMMEDIATE SOURCE:
(B) CLONE: 23H12 *RSW1* GENE

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2296..2376

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2904..3099

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3198..3370

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3594..3708

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 3824..4013

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4181..4447

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4783..5128

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 5207..5344

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 5426..5551

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 5703..5915

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 6022..6286

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 6374..6570

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 6655..7005

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 7088..8032

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTAGAAGAAG CCTGAGCCGG AGTCCTATTC AATTATCTAG AAGAAGTCTG AGCCGGAGTC 60

CCACTCGATT GTCTAGGAGA AGCCTAAGCC GGAGTCCCAT TCGATCACCT AGGAAGAGTG 120

TGAGCAGGAG TCCAGTCCGA TCATCTAGGA AGAGTGTGAG CAGAAGTCCG GTTCGTTCAT 180

CCAGGAGACG TATCAGCAGG AGTCCAGTCC GATCATCTAG GAAGAGTGTG AGCAGGAGTC 240

CTATTCGATT GTCCAGAAGA AGTATCAGCA GGAGTCCTAT TCGATTGTCC AGGAGAAGTA 300

TCAGCAGGAG TCCTGTTAGA GGAAGAAGAA GAATTAGCAG AAGTCCAGTT CCGGCAAGGA 360

GAAGGAGTGT GCGGCCTAGA TCTCCTCCTC CTGACCGCAG AAGAAGTTTG TCAAGAAGTG 420

CTTCTCCTAA TGGGCGCATA AGGAGAGGGA GAGGATTTAG CCAAAGATTC TCATACGCCC 480

GTCGATACAG AACTAGTCCA TCTCCTGATC GATCTCCTTA TCGCTTTAGT GATAGGAGTG 540

ACCGTGACAG GTGAATAGCC CACACATAAT ATAACTCCCC CTTTCTGTTA CACACTCTCG 600

TACTGAACCG TCTTTTTTAT AACGTCTTCT CTGTAGATTT AGAAGTCGCA GAAGGTTCTC 660

GCCAAGTCGG TTCAGAAGCC CACTAAGAGG AAGAACACCT CCAAGGTACT TATCCTCTTT 720

AGTACATTGT TTCAGCTGAT TCTTTACATC TAAAAGTTTC ATGAATATGG AACTAAAATT 780

GGTGATCCAA AAGAATTATT CTTGATTTCA CAACTCGAAA GTATGCTCAG GTATAGAAGA 840

AGAAGCCGCT CAGTATCGCC TGGTCTCTGT TATCGCAACC GGCGGTACAG CCGCAGCCCT 900

ATCCGTAGCC GATCTCCACC TTACAGAAAG AGAAGGTCAC CATCCGCTAG CCACAGCCTG 960

AGTCCATCGA GGTCAAGATC AAGATCAAAG TCATATTCAA AATCTCCCAT TGGGACGGGG 1020

AAAGCAAGAT CAGTGTCAAG ATCACCATCC AAGGCAAGGT CTCCATCGAA GTCGGATTCG 1080

ACATCCTCGG ATAATAGCCC AGGTGGGAAA AAGGGATTAG TAGCCTATGA TTAATGAATA 1140

ATGATTACCC TTAAGTTAAG TGTTTGTTCT TTTTACTGAG AAGAGATGGT AAAGAGAGTA 1200

AGTAGTTTAC TTCTGTAAAA CATAAGCATT GTCTTTTGCG TATGTTTGTT TGATTATGCT 1260

CCAAGATTGT TAAAAATTTC TGTTGATGTT TGCCGACATT TTTTCTTTGT TGCCATTTGC 1320

CGACAAATGT TAACTTCCAT TATTCGTTGC GGAGTTGGTT TTGGTCCAAT AATTAAACTT 1380

TCATAAAATT AAGCATAACT AAATGTGACG TTTGTCACCA AACTTTAGAA CAACGACATC 1440

GTAATTTATT TATTTGGATA ATCAATATAA TTTACGATTT CTTCCTACAT ATATATCATA 1500

TCACTATACC ACCGTCATTA TCACTATCAC TAAATATAAA AATGTTAAAA TGATTTCTTA 1560

ATGGAATTTT TTTTGTTAAA AGTTTATTGA CACAAAAAAT GAATTAAAAC TCAGAAATCT 1620

GTATACTGAA TTAAAACTTG TAAATATAAC AACAAAATGG GATTAAAAAA AGAAGTGGCA 1680

TCCATTTAAA AATTATTTGC GAATTCGCCC GTAACTTCTT AAGCTAACAA TTAGAACCTA 1740

ATCAACACTA GTTATTTTGA GTCCACCGAC AGGTGATAGC AAATAAAAAA GAACAGGCTG 1800

GTACCAGAGC CAACAACAAC GTGGCTTCTT CTTTTTTTTT TTTAATATAA TCAAACAATC 1860

ATACTTTGTC CTATCTCTTT CTTGCAATAA GATTTTGCCA CGTCACATAC TAAGAAGCTG 1920

GCGCGTCTAG TGGGGAAGCC AGAACGGCTC ACTTTAAAAA GTAGAGAGAT GATAACTTGA 1980

GCCGAATAGA GCCGAGCTGA GCTAAAACGG TGGGAGAGGA AGAGGCTACT ACTACCGTCA 2040

CCATCTCCGG TAAAATAATG TACTTGTCAT TTAAAAATTA AGAAAAAACA CATCACTCTG 2100

CGATAAAATA GGCAAAAGCA GATTTGAAGA AGAAGCAGCT TGAGATATCA AATAGAGAGA 2160

GAGAGTGACA GAGGAGTGTG TGAACATCCT TTTTTAGTAG ATTTGGGTTT TCGAGATGCC 2220

GTATTGAATC GGCTACGAAT TTCCCAATTT TGAATTTTGT GAATCTCTCT CTTTCTCTGT 2280

GTGTCGGTGG CTGCGATGGA GGCCAGTGCC GGCTTGGTTG CTGGATCCTA CCGGAGAAAC 2340

GAGCTCGTTC GGATCCGACA TGAATCTGAT GGCGGGGTCT GTTCATCTTC CCTTTTTCCC 2400

ATTTTTTTGT TATTGTTTTT CGTTCTTACA ATTTTTGATG TGTAGATCTC ATCTAGATTT 2460

CTCTGTTTCT AAATCTCGTC TCTTTTGGAT CCATAATTGG ATCATTGAAA CTCAGATTTC 2520

GCTTCCTTTG ACTGTGTAGT TAGTTAGTGT CAGTTGATCA AGTAAGTGTC TGAAAATGGA 2580

AACTTTTCTG CTCCAATTCT TCAAATTGTT GTGATCTATA TCAATTAATG CCGCATCTGT 2640

TTTCTTAAAA TCTCTTATGG AAAGTGTCGG TGGATTTCAG TTCGTTAACT TTTTTAAGCT 2700

AAAATCTTTG ACTCTTAAAG TTTAGCTTTA CTTATTGAGA TTTAGCTCAA CTAGATCTCG 2760

TTAGTTCCCG CCATGGGATA CAGACTGTGA CTCGCCTTAA TTCAGATCTG CATTGATTGT 2820

TTTGATTTAG ATCCTTGCTC ATCTCTTTCT GTAGTTTCTA ATACTCAATG ACTAACAATG 2880

ATGCAATGTT GGTCAAAGTG CAGACCAAAC CTTTGAAGAA TATGAATGGC CAGATATGTC 2940

AGATCTGTGG TGATGATGTT GGACTCGCTG AAACTGGAGA TGTCTTTGTC GCGTGTAATG 3000

AATGTGCCTT CCCTGTGTGT CGGCCTTGCT ATGAGTACGA GAGGAAAGAT GGAACTCAGT 3060

GTTGCCCTCA ATGCAAGACT AGATTCAGAC GACACAGGGG TCAGTTGTCT TTTTCTTTTT 3120

GTTGGCAATT GCTATATATG GATTTTCTCT TTTTGTTTCT TTGCTGTTGT GTTGAACAAT 3180

TTTTTGGAAT TTTCCAGGGA GTCCTCGTGT TGAAGGAGAT GAAGATGAGG ATGATGTTGA 3240

TGATATCGAG AATGAGTTCA ATTACGCCCA GGGAGCTAAC AAGGCGAGAC ACCAACGCCA 3300

TGGCGAAGAG TTTTCTTCTT CCTCTAGACA TGAATCTCAA CCAATTCCTC TTCTCACCCA 3360

TGGCCATACG GTAGGGACCT ACATTTTCCC TTTAGACTCT AGAGTGATTT GTATTACTCA 3420

ATAAATCCCT AGAGTGGTCA TTTATTACTT ACTATTCACG TTAATGTTAT ATGTGAACAA 3480

ATCTTAACAG AATTTTTTTC TGATAGTACA TGGTCATCCA AATTAAGAAA TAATAATAGA 3540

TGTTGTTAGT TGTGTCTGTT TTCAATAGAT TCATGACCTT TTTCTATACA CAGGTTTCTG 3600

GAGAGATTCG CACGCCTGAT ACACAATCTG TGCGAACTAC ATCAGGTCCT TTGGGTCCTT 3660

CTGACAGGAA TGCTATTTCA TCTCCATATA TTGATCCACG GCAACCTGGT ATTCATATGT 3720

TTTTCCCTTG TGCACGTGGT CTTTGTTAAA TGTGATTCCT ATTCATTTTT ACAACATATA 3780

TATTTTGTGT ACCGTAACTG ATAGCTCCCG CTAAAAATTG CAGTCCCTGT AAGAATCGTG 3840

GACCCGTCAA AAGACTTGAA CTCTTATGGG CTTGGTAATG TTGACTGGAA AGAAAGAGTT 3900

GAAGGCTGGA AGCTGAAGCA GGAGAAAAAT ATGTTACAGA TGACTGGTAA ATACCATGAA 3960

GGGAAAGGAG GAGAAATTGA AGGGACTGGT TCCAATGGCG AAGAACTCCA AATGTAAGTG 4020

GAAATACTAG ACCAATATCT TTATTGTCCA ACTCAAACAG CTCTTGGCCG TGATGCTAAT 4080

AACCACTCTT GGTTTCTTAT TATGTATTGA TAGACATAAT TAAGTATCTG CTTTGTTACA 4140

TTTGTTTCCT TCCACTCAAT TATGGTTCTC GTACTTACAG GGCTGATGAT ACACGTCTTC 4200

CTATGAGTCG TGTGGTGCCT ATCCCATCTT CTCGCCTAAC CCCTTATCGG GTTGTGATTA 4260

TTCTCCGGCT TATCATCTTG TGTTTCTTCT TGCAATATCG TACAACTCAC CCTGTGAAAA 4320

ATGCATATCC TTTGTGGTTG ACCTCGGTTA TCTGTGAGAT CTGGTTTGCA TTTTCTTGGC 4380

TTCTTGATCA GTTTCCCAAA TGGTACCCCA TTAACAGGGA GACTTATCTT GACCGTCTCG 4440

CTATAAGGTT GGTCTTTAAG TTTATACATC CCCTACTCTC ATCTCTCTTT TATGTATTAA 4500

CTTGATATCT TCTATCACAG TTTTCGATAG TTGACTTTTT CCCCCTGTAA ATTTAATTTA 4560

AATTTAGACA ATGGTGCATC TGAATTTTGA TTATGATATA TCTTAAGAAG ATTATGATTG 4620

TAAATCTTGA AATTTAGTAG AAAACCATCT GCAATCTACT GACCATGTGA AGTTTCCGAC 4680

TAGACTATGA TAGAAGCATG CCAAGTGGAG TGTTTATTAA GATAGAGCTT AGCTATTATA 4740

CTGATTTTAT ATGTGTTTTG ATTTTTTGGT TTCTTATTGT AGATATGATC GAGACGGTGA 4800

ACCATCACAG CTCGTTCCTG TTGATGTGTT TGTTAGTACA GTGGACCCAT TGAAAGAGCC 4860

TCCCCTTGTT ACAGCAAACA CAGTTCTCTC GATTCTTTCT GTGGACTACC CGGTAGATAA 4920

AGTAGCCTGT TATGTTTCAG ATGATGGTTC AGCTATGCTT ACCTTTGAAT CCCTTTCTGA 4980

AACCGCTGAG TTTGCAAAGA AATGGGTACC ATTTTGCAAG AAATTCAACA TTGAACCTAG 5040

GGCCCCTGAA TTCTATTTTG CCCAGAAGAT AGATTACTTG AAGGACAAGA TCCAACCGTC 5100

TTTTGTTAAA GAGCGACGAG CTATGAAGGT CATTTGAAAA GTCCACCTGC TTCTCATCCA 5160

TACGGCAAAG AGATTGACTG ACTTTTTCTT TGGTTTGTAT TGACAGAGAG AGTATGAAGA 5220

GTTTAAAGTG AGGATAAATG CTCTTGTTGC CAAAGCACAG AAAATCCCTG AAGAAGGCTG 5280

GACAATGCAG GATGGTACTC CCTGGCCTGG TAACAACACT AGAGATCATC CTGGAATGAT 5340

ACAGGTACAG TGTGGCAATC CCTTGATTGT GACAGAGAGG ATAACGTAAA GGAAACATGT 5400

TTACATCGTT TTGTTTCAAT TTCAGGTGTT CTTAGGCCAT AGTGGGGGTC TGGATACCGA 5460

TGGAAATGAG CTGCCTAGAC TCATCTATGT TTCTCGTGAA AAGCGGCCTG GATTTCAACA 5520

CCACAAAAAG GCTGGAGCTA TGAATGCATT GGTTTGTTAA CTTTCAGAAT CCTATTGTGT 5580

CCTCTATTTT ATTCTCTTGT TCACTGCCTA AGAAACGTTC TTCTTGTGTA GCCGTTGCTT 5640

CACATTCTTT TTTTTCTAGG CTATGTGTTC TCTCCTAATT TAGTATCTCT TTACTTTGAC 5700

AGATCCGTGT ATCTGCTGTT CTTACCAATG GAGCATATCT TTTGAACGTG GATTGTGATC 5760

ATTACTTTAA TAACAGTAAG GCTATTAAAG AAGCTATGTG TTTCATGATG GACCCGGCTA 5820

TTGGAAAGAA GTGCTGCTAT GTCCAGTTCC CTCAACGTTT TGACGGTATT GATTTGCACG 5880

ATCGATATGC CAACAGGAAT ATAGTCTTTT TCGATGTGAG TATCACTTCC CCATTGTCTT 5940

TTGTTTCTCT TTTGTTCATA TTTTGGTTGG ATTTACTCGT TTCTGCTATG GCCTGACTTG 6000

GATATTTGTT CTCTTGGGCA GATTAACATG AAGGGGTTGG ATGGTATCCA GGGTCCAGTA 6060

TATGTGGGTA CTGGTTGTTG TTTTAATAGG CAGGCTCTAT ATGGGTATGA TCCTGTTTTG 6120

ACGGAAGAAG ATTTAGAACC AAATATTATT GTCAAGAGCT GTTGCGGGTC AAGGAAGAAA 6180

GGTAAAAGTA GCAAGAAGTA TAACTACGAA AAGAGGAGAG GCATCAACAG AAGTGACTCC 6240

AATGCTCCAC TTTTCAATAT GGAGGACATC GATGAGGGTT TTGAAGGTTT GATTGAGCTG 6300

ATTGTGTAAT AACATCACTT CTTTATGTAA TGATTTATGT GATGGTGAAA TCTTACAATC 6360

CTTGTTTATG CAGGTTATGA TGATGAGAGG TCTATTCTAA TGTCCCAGAG GAGTGTAGAG 6420

AAGCGTTTTG GTCAGTCGCC GGTATTTATT GCGGCAACCT TCATGGAACA AGGCGGCATT 6480

CCACCAACAA CCAATCCCGC TACTCTTCTG AAGGAGGCTA TTCATGTTAT AAGCTGTGGT 6540

TACGAAGACA AGACTGAATG GGGCAAAGAG GTCAGTTTTC AAATGCAGCT ACAGAATCTT 6600

CTTATGTTCT CTTTCTTACC TGTTTGATGA CATCTTATTT GGCACTTTTG TTAGATTGGT 6660

TGGATCTATG GTTCCGTGAC GGAAGATATT CTTACTGGGT TCAAGATGCA TGCCCGGGGT 6720

TGGATATCGA TCTACTGCAA TCCTCCACGC CCTGCGTTCA AGGGATCTGC ACCAATCAAT 6780

CTTTCTGATC GTTTGAACCA AGTTCTTCGA TGGGCTTTGG GATCTATCGA GATTCTTCTT 6840

AGCAGACATT GTCCTATCTG GTATGGTTAC CATGGAAGGT TGAGACTTTT GGAGAGGATC 6900

GCTTATATCA ACACCATCGT CTATCCTATT ACATCCATCC CTCTTATTGC GTATTGTATT 6960

CTTCCCGCTT TTTGTCTCAT CACCGACAGA TTCATCATAC CCGAGGTTTG TAAAACTGAC 7020

CACACTGCTA TTTACTATTT GAATCCCATT TTGTGAATGC ATTTTTTTGT CATCATCATT 7080

GTTGCAGATA AGCAACTACG CGAGTATTTG GTTCATTCTA CTCTTCATCT CAATTGCTGT 7140

GACTGGAATC CTGGAGCTGA GATGGAGCGG TGTGAGCATT GAGGATTGGT GGAGGAACGA 7200

GCAGTTCTGG GTCATTGGTG GCACATCCGC CCATCTTTTT GCTGTCTTCC AAGGTCTACT 7260

TAAGGTTCTT GCTGGTATCG ACACCAACTT CACCGTTACA TCTAAAGCCA CAGACGAAGA 7320

TGGGGATTTT GCAGAACTCT ACATCTTCAA ATGGACAGCT CTTCTCATTC CACCAACCAC 7380

CGTCCTACTT GTGAACCTCA TAGGCATTGT GGCTGGTGTC TCTTATGCTG TAAACAGTGG 7440

CTACCAGTCG TGGGGTCCGC TTTTCGGGAA GCTCTTCTTC GCCTTATGGG TTATTGCCCA 7500

TCTCTACCCT TTCTTGAAAG GTCTGTTGGG AAGACAAAAC CGAACACCAA CCATCGTCAT 7560

TGTCTGGTCT GTTCTTCTCG CCTCCATCTT CTCGTTGCTT TGGGTCAGGA TCAATCCCTT 7620

TGTGGACGCC AATCCCAATG CCAACAACTT CAATGGCAAA GGAGGTGTCT TTTAGACCCT 7680

ATTTATATAC TTGTGTGTGC ATATATCAAA AACGCGCAAT GGGAATTCCA AATCATCTAA 7740

ACCCATCAAA CCCCAGTGAA CCGGGCAGTT AAGGTGATTC CATGTCCAAG ATTAGCTTTC 7800

TCCGAGTAGC CAGAGAAGGT GAAATTGTTC GTAACACTAT TGTAATGATT TTCCAGTGGG 7860

GAAGAAGATG TGGACCCAAA TGATACATAG TCTACAAAAA GAATTTGTTA TTCTTTCTTA 7920

TATTTATTTT ATTTAAAGCT TGTTAGACTC ACACTTATGT AATGTTGGAA CTTGTTGTCC 7980

TAAAAAGGGA TTGGAGTTTT CTTTTTATCT AAGAATCTGA AGTTTATATG CTAAGCTTTT 8040

CACTTTACTA CAAAAAGTTT ATGGATATGA TGGTGTACGT CAATTGTTGG TGCAAGTGTT 8100

GATGTCTTCG GGTGAACTCG CCCTCTTGTT TTGTCTCACC CATCAGTACA AATAGAATGA 8160

CATTTATTTT TTTGAACTTT TAACGAAATC TTTGTCATTA TGGGACTTGA TCAGTAAAGT 8220

TACATATTTG AAGAGATATT GTGTAAACTC TTATTTGAAT CAGAATCAGA TCAATCAAAA 8280

ATTGAAAACG TAAAGTTCAA ACAAAAAGGT AGAGTGAATC TTTTAATCCC CCCTCAATAC 8340

TAATTTGTGA AATCTCAAGT GGTGTAAAAT GAACCCAATT AGTATCCACA ATGTGTTTCT 8400

CTGATCAATC C 8411

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5009 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia

(vii) IMMEDIATE SOURCE:
(B) CLONE: 12C4

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 863..943

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1454..1840

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 1923..2025

(ix) FEATURE:

(A) NAME/KEY: exon
(B) LOCATION: 2122..2311

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2421..2687

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 2776..3121

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3220..3357

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3507..3623

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 3723..3935

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4027..4297

(ix) FEATURE:
(A) NAME/KEY: exon
(B) LOCATION: 4380..4576

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGGAATAAT AAGATAGGGG TTTAATGGGA GACAATCAAT CTTCAGGGGT TTTCTGGAAN 60

AACGGCGGGG TAAAAAACAA GACATCAATC GGACCCGATC ACGAGGACCC GGATCCGNAT 120

CGATAAACAG NGTAGCTTTC AATACCCCAT TTTCCCAGAA ACACCTCTCA AAAATTTTTT 180

CAAGAACTNG TATAAATATC TCAGTTTCGT TCACGCAGGT CTTTNTTATT TTGGNAANTC 240

TNTNTTCATN GTTCACCAAC TCCCTCTTGA AGGTGGGACA GAGTCCAGCT CCACCACCAC 300

CATAGCCATC GCGTCGTTTT CTCCGGGACC CACTTATTTC GTGACGTTTC TCTCTTTGTA 360

TATACATACA ATTGTTTTCA GTCTCAATTT GCTGTCCACA TTTTAACACA ACTCTATCTC 420

AGGGGTGGTG TCTGAATCTC GTCTCTCTCA TTCCTATTTA TCCCAATCTA ATCTATCACA 480

AACCCTTCCA CATTGCTTTT GTCAGTCTGT AAAATTCTCT TTGAATCAGT GAATCACTCA 540

CTTAAATCCA AAACAGTTTT TTTTTCTTTC TTTCTTTATT TGCTTGTTGT GGAATCAATA 600

GCTGTCTCCG GGAAAATTCG TTTTTTTTCT CCTTCGGGAT CTCTTTTTTT TTTTTTTTGG 660

TTTTATTTAA TAATTATCCC CGAGCCAACA TTTATTGTCG ATTCGGTTTA TTTCGTCTCC 720

TTCGTCTTCC ACTCTTACTA GTGCATGCTC TGAATCTGTA TGTAATGGGA GTTCAACAGT 780

CTGGATCCAT TATCCTAGCC GGGTCGGGTC AAGGTCTTTG AGTAAGAGAG ACAATTCGTT 840

TTGATTCGGT GTAGAAGACA TCATGAATAC TGGTGGTCGG CTCATTGCTG GCTCTCACAA 900

CAGAAACGAA TTCGTTCTCA TTAACGCCGA TGAGAGTGCC AGAGTAAGAA TAACTTTTGT 960

ANGAATTTGT GACGGAAAAA AGTTTAATTT TTTCTCTTTC TTGGGGATCT AGATTATGAG 1020

AATCTAGATG GAATATTTTG ATCTGAAATT GGAAGTTTCT AGGGAGTAAT GCCGCAACCC 1080

ACATGTTCTG TTTTTTCTTT TTTCTTTTCT TCAAGTAGTG TTGCATGATT CATACGTGTC 1140

GGCAGAGATG TCCTGAGAAC CGAATTCAAT GTTGTAGCAG TAGCAATAAG TTCAAAGAAA 1200

GTCCATTTTT TTATATTACT AATTCTGTTC TTGGTTTATT TGAGCTGGTC TTTATTGCAT 1260

TTCACCTGGA TTCAGATACT AATAACTGTC TCAATTATGT AAAAATGACA ACTTTATGAA 1320

ATTCAGTTTC ACAATTATGT AATTCATAAT CGATGAATGT TTTTCTTGAG TCTTTATCAT 1380

CTTTAGGATT TGATTAAGAT GCAATTTGAT GAAAATACTA AAAAGACTCA TGTGTTCTCA 1440

TTTCTCTATG TAGATACGAT CAGTACAAGA ACTGAGTGGG CAAACATGTC AAATCTGTGG 1500

AGATGAAATC GAATTAACGG TTAGCAGTGA GCTCTTTGTT GCTTGCAACG AATGCGCATT 1560

CCCGGTTTGT AGACCATGCT ATGAGTATGA ACGTAGAGAA GGAAATCAAG CTTGTCCTCA 1620

GTGCAAAACT CGATACAAAA GGATTAAAGG TAGTCCACGG GTTGATGGAG ATGATGAAGA 1680

AGAAGAAGAC ATTGATGATC TTGAGTATGA GTTTGATCAT GGGATGGACC CTGAACATGC 1740

CGCTGAAGCC GCACTCTCTT CACGCCTTAA CACCGGTCGT GGTGGATTGG ATTCAGCTCC 1800

ACCTGGCTCT CAGATTCCTC TTTTGACTTA TTGTGATGAA GTGAGGAATC CAAATTGTTT 1860

GTTTTCTCTG ACAATGTTGT TGCTTAGATG ATTCTTTTTC TTATTAGTCT ATGTGTTTTC 1920

AGGATGCTGA TATGTATTCT GATCGTCATG CTCTTATCGT GCCTCCTTCA ACGGGATATG 1980

GGAATCGCGT CTATCCTGCA CCGTTTACAG ATTCTTCTGC ACCTCGTATG TGTTTACTTT 2040

TATGATTCCT ACAATTTTTC TTCTTATATG ATTTGGTCAC CTTCTAATGA GTTATGAAAT 2100

GGTTTTGTTT GTTGTTTTCA GCACAGGCGA GATCAATGGT TCCTCAGAAA GATATTGCGG 2160

AATATGGTTA TGGAAGTGTT GCTTGGAAGG ACCGTATGGA AGTTTGGAAG AGACGACAAG 2220

GCGAAAAGCT TCAAGTCATT AAGCATGAAG GAGGAAACAA TGGTCGAGGT TCCAATGATG 2280

ACGACGAACT AGATGATCCT GACATGCCTA TGTAAGTTGT TAAAATCTAA CAAAAGTTCA 2340

GATGAAATGA TGCTCTGAAA TTTTGTGTTC AATGGNTTTG TTTTCTTATT GTTGTTTAAA 2400

CATTTTTCGT GCTAATTCAG GATGGATGAA GGAAGACAAC CTCTCTCAAG AAAGCTACCT 2460

ATTCGTTCAA GCAGAATAAA TCCTTACAGG ATGTTAATTC TGTGTCGCCT CGCGATTCTT 2520

GGTCTTTTCT TTCATTATAG AATTCTCCAT CCAGTCAATG ATGCATATGG ATTATGGTTA 2580

ACGTCAGTTA TATGCGAGAT ATGGTTTGCA GTGTCTTGGA TTCTTGATCA ATTCCCCAAA 2640

TGGTATCCTA TAGAACGTGA AACATACCTC GATAGACTCT CTCTCAGGTA ACATAAACCC 2700

TGAAAAGTTC TTGTCTGCAA ATATTCATTT TTTACATTCC CAAAAATTTT TGAAACTCTA 2760

TTTTTCTTAC ATAAGGTACG AGAAGGAAGG AAAACCGTCA GGATTAGCAC CTGTTGATGT 2820

TTTTGTTAGT ACAGTGGATC CGTTGAAAGA GCCACCCTTG ATTACAGCAA ACACAGTTCT 2880

TTCCATTCTA GCAGTTGATT ATCCTGTGGA TAAGGTTGCG TGTTATGTAT CAAACAATGG 2940

TGCAGCTATG CTTACATTTG AAGCTCTCTC TGATACAGCT GAGTTTGCTA GAAAATGGGT 3000

TCCTTTTTGT AAGAAGTTTA ATATCGAGCC ACGAGCTCCT GAGTGGTATT TTTCTCAGAA 3060

GATGGATTAC CTGAAGAACA AAGTTCATCC TGCTTTTGTC AGGGAACGTC GTGCTATGAA 3120

GGTTTTCTTT GCTGCTTTTT CTCTTTCTGA GTATATCCTA TCATAAAAGT GTTGTTTCAA 3180

GAATCTGATT TACGTTTTTT GCTTGTTTGT TTGTTGCAGA GAGATTATGA GGAGTTTAAA 3240

GTGAAGATAA ATGCACTGGT TGCTACTGCA CAGAAAGTGC CTGAGGAAGG TTGGACTATG 3300

CAAGATGGAA CTCCTTGGCC TGGAAACAAC GTCCGTGACC ATCCTGGAAT GATTCAGGTA 3360

ATGATGAGTT TGATTGAATA GGCAAAAAAA AAGCGGTTTT TGTCCTCTTC ACTTTGTTTC 3420

CCTGGATCTG TTAAATTGGA ATGAGCACTC TACTTCTCAA TATATCTTCA GACCGAAGCC 3480

TTTTTAAGAG ATTTTGTAAA TGACAGGTGT TCTTGGGTCA TAGTGGAGTT CGTGATACGG 3540

ATGGTAATGA GTTACCACGT CTAGTGTATG TTTCTCGTGA GAAGCGGCCT GGATTTGATC 3600

ACCACAAGAA AGCTGGAGCT ATGAATTCCT TGGTAAGTAT AATGTGTTTC TTTATTTATG 3660

AATCTCTCTT TTCGGAGCCC TGACTTCTCA TAAACTAAAA CTCATCTTAC TTCTTCTTGA 3720

AGATCCGAGT CTCTGCTGTT CTATCAAACG CTCCTTACCT TCTTAATGTC GATTGTGATC 3780

ACTACATCAA CAACAGCAAA GCAATTAGAG AATCTATGTG TTTCATGATG GACCCGCAAT 3840

CGGGAAAGAA AGTTTGTTAT GTTCAGTTTC CGCAGAGATT TGATGGGATT GATAGACATG 3900

ATAGATACTC AAACCGTAAC GTTGTGTTCT TTGATGTATG TGTCCTTATC TCTTTTGCTT 3960

TGTTTCTGTT TATGTTTTAG TGCTTTTCCT CTTTTCTCAT TTGATATTGT TTTGGTGTGG 4020

AAACAGATTA ACATGAAAGG TCTTGATGGG ATACAAGGAC CGATATATGT CGGGACAGGT 4080

TGTGTGTTTA GAAAACAGGC TCTTTATGGT TTTGATGCAC CAAAGAAGAA GAAACCACCA 4140

GGCAAAACCT GTAACTGTTG GCCTAAATGG TGTTGTTTGT GTTGTGGGTT GAGAAAGAAG 4200

AGTAAAACGA AAGCCAAAGA TAAGAAAACT AACACTAAAG AGACTTCAAA GCAGATTCAT 4260

GCGCTAGAGA ATGTCGACGA AGGTGTTATC GTCCCAGGTA AAAAAAGAAG GAAAAAAAAA 4320

ACATTTCTTA TTTGGTTTCT GTCTTGTTGA AAGTCTAAGT AGATCCTTTT GATTGTTAGT 4380

GTCAAATGTT GAGAAGAGAT CTGAAGCAAC ACAATTGAAA TTGGAGAAGA AGTTTGGACA 4440

ATCTCCGGTT TTCGTTGCCT CTGCTGTTCT ACAGAACGGT GGAGTTCCCC GTAACGCAAG 4500

CCCCGCATGT TTGTTAAGAG AAGCCATTCA AGTTATTAGC TGCGGGTACG AAGATAAAAC 4560

CGAATGGGGA AAAGAGGTAG AAAACATTAC AAAGTTTTTC AACTTCTGAA AACTAGAAAA 4620

GTTCTTGTGA TCTCATTCTT GCTGATAATC ACACGCAGAT CGGGTGGATT TATGGATCGG 4680

TGACTGAAGA TATCCTGACG GGTTTCAAGA TGCATTGCCA TGGATGGAGA TCTGTGTACT 4740

GTATGCCTAA GCGTGCAGCT TTTAAAGGAT CTGCTCCTAT TAACTTGTCA GATCGTCTTC 4800

ATCAAGTTCT ACGTTGGGCT CTTGGCTCTG TAGAGATTTT CTTGAGCAGA CATTGTCCGA 4860

TATGGTATGG TTATGGTGGT GGTTTAAAAT GGTTGGAGAG ATTCTCTTAC ATCAACTCTG 4920

TCGTCTATCC TTGGACTTCA CTTCCATTGA TCGTCTATTG TTCTCTCCCC GCGGTTTGTT 4980

TACTCACAGG AAAATTCATC GTCCCTGAG 5009

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3603 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia

(vii) IMMEDIATE SOURCE:
(B) CLONE: *RSW1* cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..3243

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG GAG GCC AGT GCC GGC TTG GTT GCT GGA TCC TAC CGG AGA AAC GAG 48
Met Glu Ala Ser Ala Gly Leu Val Ala Gly Ser Tyr Arg Arg Asn Glu
1 5 10 15

CTC GTT CGG ATC CGA CAT GAA TCT GAT GGC GGG ACC AAA CCT TTG AAG 96
Leu Val Arg Ile Arg His Glu Ser Asp Gly Gly Thr Lys Pro Leu Lys
20 25 30

```
AAT ATG AAT GGC CAG ATA TGT CAG ATC TGT GGT GAT GAT GTT GGA CTC        144
Asn Met Asn Gly Gln Ile Cys Gln Ile Cys Gly Asp Asp Val Gly Leu
         35                  40                  45

GCT GAA ACT GGA GAT GTC TTT GTC GCG TGT AAT GAA TGT GCC TTC CCT        192
Ala Glu Thr Gly Asp Val Phe Val Ala Cys Asn Glu Cys Ala Phe Pro
     50                  55                  60

GTG TGT CGG CCT TGC TAT GAG TAC GAG AGG AAA GAT GGA ACT CAG TGT        240
Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Lys Asp Gly Thr Gln Cys
 65                  70                  75                  80

TGC CCT CAA TGC AAG ACT AGA TTC AGA CGA CAC AGG GGG AGT CCT CGT        288
Cys Pro Gln Cys Lys Thr Arg Phe Arg Arg His Arg Gly Ser Pro Arg
                 85                  90                  95

GTT GAA GGA GAT GAA GAT GAG GAT GAT GTT GAT GAT ATC GAG AAT GAG        336
Val Glu Gly Asp Glu Asp Glu Asp Asp Val Asp Asp Ile Glu Asn Glu
            100                 105                 110

TTC AAT TAC GCC CAG GGA GCT AAC AAG GCG AGA CAC CAA CGC CAT GGC        384
Phe Asn Tyr Ala Gln Gly Ala Asn Lys Ala Arg His Gln Arg His Gly
        115                 120                 125

GAA GAG TTT TCT TCT TCC TCT AGA CAT GAA TCT CAA CCA ATT CCT CTT        432
Glu Glu Phe Ser Ser Ser Ser Arg His Glu Ser Gln Pro Ile Pro Leu
    130                 135                 140

CTC ACC CAT GGC CAT ACG GTT TCT GGA GAG ATT CGC ACG CCT GAT ACA        480
Leu Thr His Gly His Thr Val Ser Gly Glu Ile Arg Thr Pro Asp Thr
145                 150                 155                 160

CAA TCT GTG CGA ACT ACA TCA GGT CCT TTG GGT CCT TCT GAC AGG AAT        528
Gln Ser Val Arg Thr Thr Ser Gly Pro Leu Gly Pro Ser Asp Arg Asn
                165                 170                 175

GCT ATT TCA TCT CCA TAT ATT GAT CCA CGG CAA CCT GTC CCT GTA AGA        576
Ala Ile Ser Ser Pro Tyr Ile Asp Pro Arg Gln Pro Val Pro Val Arg
            180                 185                 190
```

ATC GTG GAC CCG TCA AAA GAC TTG AAC TCT TAT GGG CTT GGT AAT GTT 624
Ile Val Asp Pro Ser Lys Asp Leu Asn Ser Tyr Gly Leu Gly Asn Val
195 200 205

GAC TGG AAA GAA AGA GTT GAA GGC TGG AAG CTG AAG CAG GAG AAA AAT 672
Asp Trp Lys Glu Arg Val Glu Gly Trp Lys Leu Lys Gln Glu Lys Asn
210 215 220

ATG TTA CAG ATG ACT GGT AAA TAC CAT GAA GGG AAA GGA GGA GAA ATT 720
Met Leu Gln Met Thr Gly Lys Tyr His Glu Gly Lys Gly Gly Glu Ile
225 230 235 240

GAA GGG ACT GGT TCC AAT GGC GAA GAA CTC CAA ATG GCT GAT GAT ACA 768
Glu Gly Thr Gly Ser Asn Gly Glu Glu Leu Gln Met Ala Asp Asp Thr
245 250 255

CGT CTT CCT ATG AGT CGT GTG GTG CCT ATC CCA TCT TCT CGC CTA ACC 816
Arg Leu Pro Met Ser Arg Val Val Pro Ile Pro Ser Ser Arg Leu Thr
260 265 270

CCT TAT CGG GTT GTG ATT ATT CTC CGG CTT ATC ATC TTG TGT TTC TTC 864
Pro Tyr Arg Val Val Ile Ile Leu Arg Leu Ile Ile Leu Cys Phe Phe
275 280 285

TTG CAA TAT CGT ACA ACT CAC CCT GTG AAA AAT GCA TAT CCT TTG TGG 912
Leu Gln Tyr Arg Thr Thr His Pro Val Lys Asn Ala Tyr Pro Leu Trp
290 295 300

TTG ACC TCG GTT ATC TGT GAG ATC TGG TTT GCA TTT TCT TGG CTT CTT 960
Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Phe Ser Trp Leu Leu
305 310 315 320

GAT CAG TTT CCC AAA TGG TAC CCC ATT AAC AGG GAG ACT TAT CTT GAC 1008
Asp Gln Phe Pro Lys Trp Tyr Pro Ile Asn Arg Glu Thr Tyr Leu Asp
325 330 335

CGT CTC GCT ATA AGA TAT GAT CGA GAC GGT GAA CCA TCA CAG CTC GTT 1056
Arg Leu Ala Ile Arg Tyr Asp Arg Asp Gly Glu Pro Ser Gln Leu Val
340 345 350

```
CCT GTT GAT GTG TTT GTT AGT ACA GTG GAC CCA TTG AAA GAG CCT CCC      1104
Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu Pro Pro
        355                 360                 365

CTT GTT ACA GCA AAC ACA GTT CTC TCG ATT CTT TCT GTG GAC TAC CCG      1152
Leu Val Thr Ala Asn Thr Val Leu Ser Ile Leu Ser Val Asp Tyr Pro
    370                 375                 380

GTA GAT AAA GTA GCC TGT TAT GTT TCA GAT GAT GGT TCA GCT ATG CTT      1200
Val Asp Lys Val Ala Cys Tyr Val Ser Asp Asp Gly Ser Ala Met Leu
385                 390                 395                 400

ACC TTT GAA TCC CTT TCT GAA ACC GCT GAG TTT GCA AAG AAA TGG GTA      1248
Thr Phe Glu Ser Leu Ser Glu Thr Ala Glu Phe Ala Lys Lys Trp Val
                405                 410                 415

CCA TTT TGC AAG AAA TTC AAC ATT GAA CCT AGG GCC CCT GAA TTC TAT      1296
Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro Glu Phe Tyr
            420                 425                 430

TTT GCC CAG AAG ATA GAT TAC TTG AAG GAC AAG ATC CAA CCG TCT TTT      1344
Phe Ala Gln Lys Ile Asp Tyr Leu Lys Asp Lys Ile Gln Pro Ser Phe
        435                 440                 445

GTT AAA GAG CGA CGA GCT ATG AAG AGA GAG TAT GAA GAG TTT AAA GTG      1392
Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val
    450                 455                 460

AGG ATA AAT GCT CTT GTT GCC AAA GCA CAG AAA ATC CCT GAA GAA GGC      1440
Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Ile Pro Glu Glu Gly
465                 470                 475                 480

TGG ACA ATG CAG GAT GGT ACT CCC TGG CCT GGT AAC AAC ACT AGA GAT      1488
Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp
                485                 490                 495

CAT CCT GGA ATG ATA CAG GTG TTC TTA GGC CAT AGT GGG GGT CTG GAT      1536
His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp
            500                 505                 510
```

ACC GAT GGA AAT GAG CTG CCT AGA CTC ATC TAT GTT TCT CGT GAA AAG 1584
Thr Asp Gly Asn Glu Leu Pro Arg Leu Ile Tyr Val Ser Arg Glu Lys
515 520 525

CGG CCT GGA TTT CAA CAC CAC AAA AAG GCT GGA GCT ATG AAT GCA TTG 1632
Arg Pro Gly Phe Gln His His Lys Lys Ala Gly Ala Met Asn Ala Leu
530 535 540

ATC CGT GTA TCT GCT GTT CTT ACC AAT GGA GCA TAT CTT TTG AAC GTG 1680
Ile Arg Val Ser Ala Val Leu Thr Asn Gly Ala Tyr Leu Leu Asn Val
545 550 555 560

GAT TGT GAT CAT TAC TTT AAT AAC AGT AAG GCT ATT AAA GAA GCT ATG 1728
Asp Cys Asp His Tyr Phe Asn Asn Ser Lys Ala Ile Lys Glu Ala Met
565 570 575

TGT TTC ATG ATG GAC CCG GCT ATT GGA AAG AAG TGC TGC TAT GTC CAG 1776
Cys Phe Met Met Asp Pro Ala Ile Gly Lys Lys Cys Cys Tyr Val Gln
580 585 590

TTC CCT CAA CGT TTT GAC GGT ATT GAT TTG CAC GAT CGA TAT GCC AAC 1824
Phe Pro Gln Arg Phe Asp Gly Ile Asp Leu His Asp Arg Tyr Ala Asn
595 600 605

AGG AAT ATA GTC TTT TTC GAT ATT AAC ATG AAG GGG TTG GAT GGT ATC 1872
Arg Asn Ile Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Ile
610 615 620

CAG GGT CCA GTA TAT GTG GGT ACT GGT TGT TGT TTT AAT AGG CAG GCT 1920
Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Cys Phe Asn Arg Gln Ala
625 630 635 640

CTA TAT GGG TAT GAT CCT GTT TTG ACG GAA GAA GAT TTA GAA CCA AAT 1968
Leu Tyr Gly Tyr Asp Pro Val Leu Thr Glu Glu Asp Leu Glu Pro Asn
645 650 655

ATT ATT GTC AAG AGC TGT TGC GGG TCA AGG AAG AAA GGT AAA AGT AGC 2016
Ile Ile Val Lys Ser Cys Cys Gly Ser Arg Lys Lys Gly Lys Ser Ser
660 665 670

```
AAG AAG TAT AAC TAC GAA AAG AGG AGA GGC ATC AAC AGA AGT GAC TCC      2064
Lys Lys Tyr Asn Tyr Glu Lys Arg Arg Gly Ile Asn Arg Ser Asp Ser
        675                 680                 685

AAT GCT CCA CTT TTC AAT ATG GAG GAC ATC GAT GAG GGT TTT GAA GGT      2112
Asn Ala Pro Leu Phe Asn Met Glu Asp Ile Asp Glu Gly Phe Glu Gly
    690                 695                 700

TAT GAT GAT GAG AGG TCT ATT CTA ATG TCC CAG AGG AGT GTA GAG AAG      2160
Tyr Asp Asp Glu Arg Ser Ile Leu Met Ser Gln Arg Ser Val Glu Lys
705                 710                 715                 720

CGT TTT GGT CAG TCG CCG GTA TTT ATT GCG GCA ACC TTC ATG GAA CAA      2208
Arg Phe Gly Gln Ser Pro Val Phe Ile Ala Ala Thr Phe Met Glu Gln
                725                 730                 735

GGC GGC ATT CCA CCA ACA ACC AAT CCC GCT ACT CTT CTG AAG GAG GCT      2256
Gly Gly Ile Pro Pro Thr Thr Asn Pro Ala Thr Leu Leu Lys Glu Ala
            740                 745                 750

ATT CAT GTT ATA AGC TGT GGT TAC GAA GAC AAG ACT GAA TGG GGC AAA      2304
Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp Gly Lys
        755                 760                 765

GAG ATT GGT TGG ATC TAT GGT TCC GTG ACG GAA GAT ATT CTT ACT GGG      2352
Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly
    770                 775                 780

TTC AAG ATG CAT GCC CGG GGT TGG ATA TCG ATC TAC TGC AAT CCT CCA      2400
Phe Lys Met His Ala Arg Gly Trp Ile Ser Ile Tyr Cys Asn Pro Pro
785                 790                 795                 800

CGC CCT GCG TTC AAG GGA TCT GCA CCA ATC AAT CTT TCT GAT CGT TTG      2448
Arg Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu
                805                 810                 815

AAC CAA GTT CTT CGA TGG GCT TTG GGA TCT ATC GAG ATT CTT CTT AGC      2496
Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Ile Glu Ile Leu Leu Ser
            820                 825                 830
```

```
AGA CAT TGT CCT ATC TGG TAT GGT TAC CAT GGA AGG TTG AGA CTT TTG      2544
Arg His Cys Pro Ile Trp Tyr Gly Tyr His Gly Arg Leu Arg Leu Leu
        835                 840                 845

GAG AGG ATC GCT TAT ATC AAC ACC ATC GTC TAT CCT ATT ACA TCC ATC      2592
Glu Arg Ile Ala Tyr Ile Asn Thr Ile Val Tyr Pro Ile Thr Ser Ile
    850                 855                 860

CCT CTT ATT GCG TAT TGT ATT CTT CCC GCT TTT TGT CTC ATC ACC GAC      2640
Pro Leu Ile Ala Tyr Cys Ile Leu Pro Ala Phe Cys Leu Ile Thr Asp
865                 870                 875                 880

AGA TTC ATC ATA CCC GAG ATA AGC AAC TAC GCG AGT ATT TGG TTC ATT      2688
Arg Phe Ile Ile Pro Glu Ile Ser Asn Tyr Ala Ser Ile Trp Phe Ile
                885                 890                 895

CTA CTC TTC ATC TCA ATT GCT GTG ACT GGA ATC CTG GAG CTG AGA TGG      2736
Leu Leu Phe Ile Ser Ile Ala Val Thr Gly Ile Leu Glu Leu Arg Trp
            900                 905                 910

AGC GGT GTG AGC ATT GAG GAT TGG TGG AGG AAC GAG CAG TTC TGG GTC      2784
Ser Gly Val Ser Ile Glu Asp Trp Trp Arg Asn Glu Gln Phe Trp Val
        915                 920                 925

ATT GGT GGC ACA TCC GCC CAT CTT TTT GCT GTC TTC CAA GGT CTA CTT      2832
Ile Gly Gly Thr Ser Ala His Leu Phe Ala Val Phe Gln Gly Leu Leu
    930                 935                 940

AAG GTT CTT GCT GGT ATC GAC ACC AAC TTC ACC GTT ACA TCT AAA GCC      2880
Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr Ser Lys Ala
945                 950                 955                 960

ACA GAC GAA GAT GGG GAT TTT GCA GAA CTC TAC ATC TTC AAA TGG ACA      2928
Thr Asp Glu Asp Gly Asp Phe Ala Glu Leu Tyr Ile Phe Lys Trp Thr
                965                 970                 975

GCT CTT CTC ATT CCA CCA ACC ACC GTC CTA CTT GTG AAC CTC ATA GGC      2976
Ala Leu Leu Ile Pro Pro Thr Thr Val Leu Leu Val Asn Leu Ile Gly
            980                 985                 990
```

```
ATT GTG GCT GGT GTC TCT TAT GCT GTA AAC AGT GGC TAC CAG TCG TGG      3024
Ile Val Ala Gly Val Ser Tyr Ala Val Asn Ser Gly Tyr Gln Ser Trp
        995                 1000                1005

GGT CCG CTT TTC GGG AAG CTC TTC TTC GCC TTA TGG GTT ATT GCC CAT      3072
Gly Pro Leu Phe Gly Lys Leu Phe Phe Ala Leu Trp Val Ile Ala His
    1010                1015                1020

CTC TAC CCT TTC TTG AAA GGT CTG TTG GGA AGA CAA AAC CGA ACA CCA      3120
Leu Tyr Pro Phe Leu Lys Gly Leu Leu Gly Arg Gln Asn Arg Thr Pro
1025                1030                1035                1040

ACC ATC GTC ATT GTC TGG TCT GTT CTT CTC GCC TCC ATC TTC TCG TTG      3168
Thr Ile Val Ile Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu
                1045                1050                1055

CTT TGG GTC AGG ATC AAT CCC TTT GTG GAC GCC AAT CCC AAT GCC AAC      3216
Leu Trp Val Arg Ile Asn Pro Phe Val Asp Ala Asn Pro Asn Ala Asn
            1060                1065                1070

AAC TTC AAT GGC AAA GGA GGT GTC TTT TAGACCCTAT TTATATACTT            3263
Asn Phe Asn Gly Lys Gly Gly Val Phe
        1075                1080

GTGTGTGCAT ATATCAAAAA CGCGCAATGG GAATTCCAAA TCATCTAAAC CCATCAAACC    3323

CCAGTGAACC GGGCAGTTAA GGTGATTCCA TGTCCAAGAT TAGCTTTCTC CGAGTAGCCA    3383

GAGAAGGTGA AATTGTTCGT AACACTATTG TAATGATTTT CCAGTGGGGA AGAAGATGTG    3443

GACCCAAATG ATACATAGTC TACAAAAAGA ATTTGTTATT CTTTCTTATA TTTATTTTAT    3503

TTAAAGCTTG TTAGACTCAC ACTTATGTAA TGTTGGAACT TGTTGTCCTA AAAAGGGATT    3563

GGAGTTTTCT TTTTATCTAA GAATCTGAAG TTTATATGCT                          3603
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1081 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Glu Ala Ser Ala Gly Leu Val Ala Gly Ser Tyr Arg Arg Asn Glu
 1               5                  10                  15

Leu Val Arg Ile Arg His Glu Ser Asp Gly Gly Thr Lys Pro Leu Lys
            20                  25                  30

Asn Met Asn Gly Gln Ile Cys Gln Ile Cys Gly Asp Asp Val Gly Leu
        35                  40                  45

Ala Glu Thr Gly Asp Val Phe Val Ala Cys Asn Glu Cys Ala Phe Pro
    50                  55                  60

Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Lys Asp Gly Thr Gln Cys
65                  70                  75                  80

Cys Pro Gln Cys Lys Thr Arg Phe Arg Arg His Arg Gly Ser Pro Arg
                85                  90                  95

Val Glu Gly Asp Glu Asp Glu Asp Asp Val Asp Asp Ile Glu Asn Glu
            100                 105                 110

Phe Asn Tyr Ala Gln Gly Ala Asn Lys Ala Arg His Gln Arg His Gly
        115                 120                 125

Glu Glu Phe Ser Ser Ser Ser Arg His Glu Ser Gln Pro Ile Pro Leu
    130                 135                 140

Leu Thr His Gly His Thr Val Ser Gly Glu Ile Arg Thr Pro Asp Thr
145                 150                 155                 160

Gln Ser Val Arg Thr Thr Ser Gly Pro Leu Gly Pro Ser Asp Arg Asn
                165                 170                 175
```

```
Ala Ile Ser Ser Pro Tyr Ile Asp Pro Arg Gln Pro Val Pro Val Arg
            180                 185                 190

Ile Val Asp Pro Ser Lys Asp Leu Asn Ser Tyr Gly Leu Gly Asn Val
        195                 200                 205

Asp Trp Lys Glu Arg Val Glu Gly Trp Lys Leu Lys Gln Glu Lys Asn
    210                 215                 220

Met Leu Gln Met Thr Gly Lys Tyr His Glu Gly Lys Gly Gly Glu Ile
225                 230                 235                 240

Glu Gly Thr Gly Ser Asn Gly Glu Glu Leu Gln Met Ala Asp Asp Thr
                245                 250                 255

Arg Leu Pro Met Ser Arg Val Val Pro Ile Pro Ser Ser Arg Leu Thr
            260                 265                 270

Pro Tyr Arg Val Val Ile Ile Leu Arg Leu Ile Ile Leu Cys Phe Phe
        275                 280                 285

Leu Gln Tyr Arg Thr Thr His Pro Val Lys Asn Ala Tyr Pro Leu Trp
    290                 295                 300

Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Phe Ser Trp Leu Leu
305                 310                 315                 320

Asp Gln Phe Pro Lys Trp Tyr Pro Ile Asn Arg Glu Thr Tyr Leu Asp
                325                 330                 335

Arg Leu Ala Ile Arg Tyr Asp Arg Asp Gly Glu Pro Ser Gln Leu Val
            340                 345                 350

Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu Pro Pro
        355                 360                 365

Leu Val Thr Ala Asn Thr Val Leu Ser Ile Leu Ser Val Asp Tyr Pro
    370                 375                 380
```

```
Val Asp Lys Val Ala Cys Tyr Val Ser Asp Asp Gly Ser Ala Met Leu
385                 390                 395                 400

Thr Phe Glu Ser Leu Ser Glu Thr Ala Glu Phe Ala Lys Lys Trp Val
                405                 410                 415

Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro Glu Phe Tyr
            420                 425                 430

Phe Ala Gln Lys Ile Asp Tyr Leu Lys Asp Lys Ile Gln Pro Ser Phe
        435                 440                 445

Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val
    450                 455                 460

Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Ile Pro Glu Glu Gly
465                 470                 475                 480

Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp
                485                 490                 495

His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp
            500                 505                 510

Thr Asp Gly Asn Glu Leu Pro Arg Leu Ile Tyr Val Ser Arg Glu Lys
        515                 520                 525

Arg Pro Gly Phe Gln His His Lys Lys Ala Gly Ala Met Asn Ala Leu
    530                 535                 540

Ile Arg Val Ser Ala Val Leu Thr Asn Gly Ala Tyr Leu Leu Asn Val
545                 550                 555                 560

Asp Cys Asp His Tyr Phe Asn Asn Ser Lys Ala Ile Lys Glu Ala Met
                565                 570                 575

Cys Phe Met Met Asp Pro Ala Ile Gly Lys Lys Cys Cys Tyr Val Gln
            580                 585                 590
```

```
Phe Pro Gln Arg Phe Asp Gly Ile Asp Leu His Asp Arg Tyr Ala Asn
        595                 600                 605

Arg Asn Ile Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Ile
    610                 615                 620

Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Cys Phe Asn Arg Gln Ala
625                 630                 635                 640

Leu Tyr Gly Tyr Asp Pro Val Leu Thr Glu Glu Asp Leu Glu Pro Asn
                645                 650                 655

Ile Ile Val Lys Ser Cys Cys Gly Ser Arg Lys Lys Gly Lys Ser Ser
            660                 665                 670

Lys Lys Tyr Asn Tyr Glu Lys Arg Arg Gly Ile Asn Arg Ser Asp Ser
        675                 680                 685

Asn Ala Pro Leu Phe Asn Met Glu Asp Ile Asp Glu Gly Phe Glu Gly
    690                 695                 700

Tyr Asp Asp Glu Arg Ser Ile Leu Met Ser Gln Arg Ser Val Glu Lys
705                 710                 715                 720

Arg Phe Gly Gln Ser Pro Val Phe Ile Ala Ala Thr Phe Met Glu Gln
                725                 730                 735

Gly Gly Ile Pro Pro Thr Thr Asn Pro Ala Thr Leu Leu Lys Glu Ala
            740                 745                 750

Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp Gly Lys
        755                 760                 765

Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly
    770                 775                 780

Phe Lys Met His Ala Arg Gly Trp Ile Ser Ile Tyr Cys Asn Pro Pro
785                 790                 795                 800
```

Arg Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu
805 810 815

Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Ile Glu Ile Leu Leu Ser
820 825 830

Arg His Cys Pro Ile Trp Tyr Gly Tyr His Gly Arg Leu Arg Leu Leu
835 840 845

Glu Arg Ile Ala Tyr Ile Asn Thr Ile Val Tyr Pro Ile Thr Ser Ile
850 855 860

Pro Leu Ile Ala Tyr Cys Ile Leu Pro Ala Phe Cys Leu Ile Thr Asp
865 870 875 880

Arg Phe Ile Ile Pro Glu Ile Ser Asn Tyr Ala Ser Ile Trp Phe Ile
885 890 895

Leu Leu Phe Ile Ser Ile Ala Val Thr Gly Ile Leu Glu Leu Arg Trp
900 905 910

Ser Gly Val Ser Ile Glu Asp Trp Trp Arg Asn Glu Gln Phe Trp Val
915 920 925

Ile Gly Gly Thr Ser Ala His Leu Phe Ala Val Phe Gln Gly Leu Leu
930 935 940

Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr Ser Lys Ala
945 950 955 960

Thr Asp Glu Asp Gly Asp Phe Ala Glu Leu Tyr Ile Phe Lys Trp Thr
965 970 975

Ala Leu Leu Ile Pro Pro Thr Thr Val Leu Leu Val Asn Leu Ile Gly
980 985 990

Ile Val Ala Gly Val Ser Tyr Ala Val Asn Ser Gly Tyr Gln Ser Trp
995 1000 1005

Gly Pro Leu Phe Gly Lys Leu Phe Phe Ala Leu Trp Val Ile Ala His
1010 1015 1020

Leu Tyr Pro Phe Leu Lys Gly Leu Leu Gly Arg Gln Asn Arg Thr Pro
1025 1030 1035 1040

Thr Ile Val Ile Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu
1045 1050 1055

Leu Trp Val Arg Ile Asn Pro Phe Val Asp Ala Asn Pro Asn Ala Asn
1060 1065 1070

Asn Phe Asn Gly Lys Gly Gly Val Phe
1075 1080

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia

(vii) IMMEDIATE SOURCE:
(B) CLONE: Ath-A

(ix) FEATURE:
(A) NAME/KEY: CDS

(B) LOCATION: 239..3490

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
GTCGACACTA AGTGGATCCA AGAATTCGC GGCCGCGTCG ATACGGCTGC GAGAAGACGA      60

CAGAAGGGGA TTGTCGATTC GGTTTATTTC GTCTCCTTCG TCTTCCACTC TTACTAGTGC     120

ATGCTCTGAA TCTGTATGTA ATGGGAGTTC AACAGTCTGG ATCCATTATC CTAGCCGGGT     180

CGGGTCAAGG TCTTTGAATA AGAGAGACAA TTCGTTTTGA TTCGGTGTAG AAGACATC       238

ATG AAT ACT GGT GGT CGG CTC ATT GCT GGC TCT CAC AAC AGA AAC GAA         286
Met Asn Thr Gly Gly Arg Leu Ile Ala Gly Ser His Asn Arg Asn Glu
 1               5                  10                  15

TTC GTT CTC ATT AAC GCC GAT GAG AGT GCC AGA ATA CGA TCA GTA CAA        334
Phe Val Leu Ile Asn Ala Asp Glu Ser Ala Arg Ile Arg Ser Val Gln
            20                  25                  30

GAA CTG AGT GGG CAA ACA TGT CAA ATC TGT GGA GAT GAA ATC GAA TTA        382
Glu Leu Ser Gly Gln Thr Cys Gln Ile Cys Gly Asp Glu Ile Glu Leu
        35                  40                  45

ACG GTT AGC AGT GAG CTC TTT GTT GCT TGC AAC GAA TGC GCA TTC CCG        430
Thr Val Ser Ser Glu Leu Phe Val Ala Cys Asn Glu Cys Ala Phe Pro
    50                  55                  60

GTT TGT AGA CCA TGC TAT GAG TAT GAA CGT AGA GAA GGA AAT CAA GCT        478
Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Arg Glu Gly Asn Gln Ala
65                  70                  75                  80

TGT CCT CAG TGC AAA ACT CGA TAC AAA AGG ATT AAA GGT AGT CCA CGG        526
Cys Pro Gln Cys Lys Thr Arg Tyr Lys Arg Ile Lys Gly Ser Pro Arg
                85                  90                  95

GTT GAT GGA GAT GAT GAA GAA GAA GAA GAC ATT GAT GAT CTT GAG TAT        574
Val Asp Gly Asp Asp Glu Glu Glu Glu Asp Ile Asp Asp Leu Glu Tyr
            100                 105                 110
```

```
GAG TTT GAT CAT GGG ATG GAC CCT GAA CAT GCC GCT GAA GCC GCA CTC      622
Glu Phe Asp His Gly Met Asp Pro Glu His Ala Ala Glu Ala Ala Leu
        115                 120                 125

TCT TCA CGC CTT AAC ACC GGT CGT GGT GGA TTG GAT TCA GCT CCA CCT      670
Ser Ser Arg Leu Asn Thr Gly Arg Gly Gly Leu Asp Ser Ala Pro Pro
    130                 135                 140

GGC TCT CAG ATT CCT CTT TTG ACT TAT TGT GAT GAA GAT GCT GAT ATG      718
Gly Ser Gln Ile Pro Leu Leu Thr Tyr Cys Asp Glu Asp Ala Asp Met
145                 150                 155                 160

TAT TCT GAT CGT CAT GCT CTT ATC GTG CCT CCT TCA ACG GGA TAT GGG      766
Tyr Ser Asp Arg His Ala Leu Ile Val Pro Pro Ser Thr Gly Tyr Gly
                165                 170                 175

AAT CGC GTC TAT CCT GCA CCG TTT ACA GAT TCT TCT GCA CCT CCA CAG      814
Asn Arg Val Tyr Pro Ala Pro Phe Thr Asp Ser Ser Ala Pro Pro Gln
            180                 185                 190

GCG AGA TCA ATG GTT CCT CAG AAA GAT ATT GCG GAA TAT GGT TAT GGA      862
Ala Arg Ser Met Val Pro Gln Lys Asp Ile Ala Glu Tyr Gly Tyr Gly
        195                 200                 205

AGT GTT GCT TGG AAG GAC CGT ATG GAA GTT TGG AAG AGA CGA CAA GGC      910
Ser Val Ala Trp Lys Asp Arg Met Glu Val Trp Lys Arg Arg Gln Gly
    210                 215                 220

GAA AAG CTT CAA GTC ATT AAG CAT GAA GGA GGA AAC AAT GGT CGA GGT      958
Glu Lys Leu Gln Val Ile Lys His Glu Gly Gly Asn Asn Gly Arg Gly
225                 230                 235                 240

TCC AAT GAT GAC GAC GAA CTA GAT GAT CCT GAC ATG CCT ATG ATG GAT     1006
Ser Asn Asp Asp Asp Glu Leu Asp Asp Pro Asp Met Pro Met Met Asp
                245                 250                 255

GAA GGA AGA CAA CCT CTC TCA AGA AAG CTA CCT ATT CGT TCA AGC AGA     1054
Glu Gly Arg Gln Pro Leu Ser Arg Lys Leu Pro Ile Arg Ser Ser Arg
            260                 265                 270
```

```
ATA AAT CCT TAC AGG ATG TTA ATT CTG TGT CGC CTC GCG ATT CTT GGT      1102
Ile Asn Pro Tyr Arg Met Leu Ile Leu Cys Arg Leu Ala Ile Leu Gly
        275                 280                 285

CTT TTC TTT CAT TAT AGA ATT CTC CAT CCA GTC AAT GAT GCA TAT GGA      1150
Leu Phe Phe His Tyr Arg Ile Leu His Pro Val Asn Asp Ala Tyr Gly
    290                 295                 300

TTA TGG TTA ACG TCA GTT ATA TGC GAA ATA TGG TTT GCA GTG TCT TGG      1198
Leu Trp Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Val Ser Trp
305                 310                 315                 320

ATT CTT GAT CAA TTC CCC AAA TGG TAT CCT ATA GAA CGT GAA ACA TAC      1246
Ile Leu Asp Gln Phe Pro Lys Trp Tyr Pro Ile Glu Arg Glu Thr Tyr
                325                 330                 335

CTC GAT AGA CTC TCT CTC AGG TAC GAG AAG GAA GGA AAA CCG TCA GGA      1294
Leu Asp Arg Leu Ser Leu Arg Tyr Glu Lys Glu Gly Lys Pro Ser Gly
            340                 345                 350

TTA GCA CCT GTT GAT GTT TTT GTT AGT ACA GTG GAT CCG TTG AAA GAG      1342
Leu Ala Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu
        355                 360                 365

CCC CCC TTG ATT ACA GCA AAC ACA GTT CTT TCC ATT CTA GCA GTT GAT      1390
Pro Pro Leu Ile Thr Ala Asn Thr Val Leu Ser Ile Leu Ala Val Asp
    370                 375                 380

TAT CCT GTG GAT AAG GTT GCG TGT TAT GTA TCA AAC AAT GGT GCA GCT      1438
Tyr Pro Val Asp Lys Val Ala Cys Tyr Val Ser Asn Asn Gly Ala Ala
385                 390                 395                 400

ATG CTT ACA TTT GAA GCT CTC TCT GAT ACA GCT GAT TTT GCT ACA AAA      1486
Met Leu Thr Phe Glu Ala Leu Ser Asp Thr Ala Asp Phe Ala Thr Lys
                405                 410                 415

TGG GTT CCT TTT TGT AAG AAG TTT AAT ATC GAG CCA CGA GCT CCT GAG      1534
Trp Val Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro Glu
            420                 425                 430
```

TGG TAT TTT TCT CAG AAG ATG GAT TAC CTG AAG AAC AAA GTT CAT CCT 1582
Trp Tyr Phe Ser Gln Lys Met Asp Tyr Leu Lys Asn Lys Val His Pro
435 440 445

GCT TTT GTC AGG GAA CGT CGT GCT ATG AAG AGA GAT TAT GAA GAG TTT 1630
Ala Phe Val Arg Glu Arg Arg Ala Met Lys Arg Asp Tyr Glu Glu Phe
450 455 460

AAA GTG AAG ATA AAT GCA CTG GTT GCT ACT GCA CAG AAA GTG CCT GAG 1678
Lys Val Lys Ile Asn Ala Leu Val Ala Thr Ala Gln Lys Val Pro Glu
465 470 475 480

GAA CGT TGG ACT ATG CAA GAT GGA ACT CCT TGG CCT GGA AAC AAC GTC 1726
Glu Arg Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Val
485 490 495

CGT GAC CAT CCT GGA ATG ATT CAG GTG TTC TTG GGT CAT AGT GGA GTT 1774
Arg Asp His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly Val
500 505 510

CGT GAT ACG GAT GGT AAT GAG TTA CCA CGT CTA GTG TAT GTT TCT CGT 1822
Arg Asp Thr Asp Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg
515 520 525

GAG AAG CGG CCT GGA TTT GAT CAC CAC AAG AAA GCT GGA GCT ATG AAT 1870
Glu Lys Arg Pro Gly Phe Asp His His Lys Lys Ala Gly Ala Met Asn
530 535 540

TCC TTG ATC CGA GTC TCT GCT GTT CTA TCA AAC GCT CCT TAC CTT CTT 1918
Ser Leu Ile Arg Val Ser Ala Val Leu Ser Asn Ala Pro Tyr Leu Leu
545 550 555 560

AAT GTC GAT TGT GAT CAC TAC ATC AAC AAC AGC AAA GCA ATT AGA GAA 1966
Asn Val Asp Cys Asp His Tyr Ile Asn Asn Ser Lys Ala Ile Arg Glu
565 570 575

TCT ATG TGT TTC ATG ATG GAC CCG CAA TCG GGA AAG AAA GTT TGT TAT 2014
Ser Met Cys Phe Met Met Asp Pro Gln Ser Gly Lys Lys Val Cys Tyr
580 585 590

```
GTT CAG TTT CCG CAG AGA TTT GAT GGG ATT GAT AGA CAT GAT AGA TAC      2062
Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Arg His Asp Arg Tyr
        595                 600                 605

TCA AAC CGT AAC GTT GTG TTC TTT GAT ATT AAC ATG AAA GGT CTT GAT      2110
Ser Asn Arg Asn Val Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp
    610                 615                 620

GGG ATA CAA GGA CCG ATA TAT GTC GGG ACA GGT TGT GTG TTT AGA AAA      2158
Gly Ile Gln Gly Pro Ile Tyr Val Gly Thr Gly Cys Val Phe Arg Lys
625                 630                 635                 640

CAG GCT CTT TAT GGT TTT GAT GCA CCA AAG AAG AAG AAA CCA CCA GGC      2206
Gln Ala Leu Tyr Gly Phe Asp Ala Pro Lys Lys Lys Lys Pro Pro Gly
                645                 650                 655

AAA ACC TGT AAC TGT TGG CCT AAA TGG TGT TGT TTG TGT TGT GGG TTG      2254
Lys Thr Cys Asn Cys Trp Pro Lys Trp Cys Cys Leu Cys Cys Gly Leu
            660                 665                 670

AGA AAG AAG AGT AAA ACG AAA GCC ACA GAT AAG AAA ACT AAC ACT AAA      2302
Arg Lys Lys Ser Lys Thr Lys Ala Thr Asp Lys Lys Thr Asn Thr Lys
        675                 680                 685

GAG ACT TCA AAG CAG ATT CAT GCG CTA GAG AAT GTC GAC GAA GGT GTT      2350
Glu Thr Ser Lys Gln Ile His Ala Leu Glu Asn Val Asp Glu Gly Val
    690                 695                 700

ATC GTC CCA GTG TCA AAT GTT GAG AAG AGA TCT GAA GCA ACA CAA TTG      2398
Ile Val Pro Val Ser Asn Val Glu Lys Arg Ser Glu Ala Thr Gln Leu
705                 710                 715                 720

AAA TTG GAG AAG AAG TTT GGA CAA TCT CCG GTT TTC GTT GCC TCT GCT      2446
Lys Leu Glu Lys Lys Phe Gly Gln Ser Pro Val Phe Val Ala Ser Ala
                725                 730                 735

GTT CTA CAG AAC GGT GGA GTT CCC CGT AAC GCA AGC CCC GCA TGT TTG      2494
Val Leu Gln Asn Gly Gly Val Pro Arg Asn Ala Ser Pro Ala Cys Leu
            740                 745                 750
```

```
TTA AGA GAA GCC ATT CAA GTT ATT AGC TGC GGG TAC CAA GAT AAA ACC      2542
Leu Arg Glu Ala Ile Gln Val Ile Ser Cys Gly Tyr Gln Asp Lys Thr
        755                 760                 765

GAA TGG GGA AAA GAG ATC GGG TGG ATT TAT GGA TCG GTG ACT GAA GAT      2590
Glu Trp Gly Lys Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp
    770                 775                 780

ATC CTG ACG GGT TTC AAG ATG CAT TGC CAT GGA TGG AGA TCT GTG TAC      2638
Ile Leu Thr Gly Phe Lys Met His Cys His Gly Trp Arg Ser Val Tyr
785                 790                 795                 800

TGT ATG CCT AAG CGT GCA GCT TTT AAA GGA TCT GCT CCT ATT AAC TTG      2686
Cys Met Pro Lys Arg Ala Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu
                805                 810                 815

TCA GAT CGT CTT CAT CAA GTT CTA CGT TGG GCT CTT GGC TCT GTA GAG      2734
Ser Asp Arg Leu His Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu
            820                 825                 830

ATT TTC TTG AGC AGA CAT TGT CCG ATA TGG TAT GGT TAT GGT GGT GGT      2782
Ile Phe Leu Ser Arg His Cys Pro Ile Trp Tyr Gly Tyr Gly Gly Gly
        835                 840                 845

TTA AAA TGG TTG GAG AGA TTC TCT TAC ATC AAC TCT GTC GTC TAT CCT      2830
Leu Lys Trp Leu Glu Arg Phe Ser Tyr Ile Asn Ser Val Val Tyr Pro
    850                 855                 860

TGG ACT TCA CTT CCA TTG ATC GTC TAT TGT TCT CTC CCC GCG GTT TGT      2878
Trp Thr Ser Leu Pro Leu Ile Val Tyr Cys Ser Leu Pro Ala Val Cys
865                 870                 875                 880

TTA CTC ACA GGA AAA TTC ATC GTC CCT GAG ATA AGC AAC TAC GCA GGT      2926
Leu Leu Thr Gly Lys Phe Ile Val Pro Glu Ile Ser Asn Tyr Ala Gly
                885                 890                 895

ATA CTC TTC ATG CTC ATG TTC ATA TCC ATA GCA GTA ACT GGA ATC CTC      2974
Ile Leu Phe Met Leu Met Phe Ile Ser Ile Ala Val Thr Gly Ile Leu
            900                 905                 910
```

```
GAA ATG CAA TGG GGA GGT GTC GGA ATC GAT GAT TGG TGG AGA AAC GAG      3022
Glu Met Gln Trp Gly Gly Val Gly Ile Asp Asp Trp Trp Arg Asn Glu
        915                 920                 925

CAG TTT TGG GTA ATC GGA GGG GCC TCC TCG CAT CTA TTT GCT CTG TTT      3070
Gln Phe Trp Val Ile Gly Gly Ala Ser Ser His Leu Phe Ala Leu Phe
    930                 935                 940

CAA GGT TTG CTC AAA GTT CTA GCC GGA GTT AAC ACG AAT TTC ACA GTC      3118
Gln Gly Leu Leu Lys Val Leu Ala Gly Val Asn Thr Asn Phe Thr Val
945                 950                 955                 960

ACT TCA AAA GCA GCA GAC GAT GGA GCT TTC TCT GAG CTT TAC ATC TTC      3166
Thr Ser Lys Ala Ala Asp Asp Gly Ala Phe Ser Glu Leu Tyr Ile Phe
                965                 970                 975

AAG TGG ACA ACT TTG TTG ATT CCT CCG ACA ACA CTT CTG ATC ATT AAC      3214
Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Leu Leu Ile Ile Asn
            980                 985                 990

ATC ATT GGA GTT ATT GTC GGC GTT TCT GAT GCC ATT AGC AAT GGC TAT      3262
Ile Ile Gly Val Ile Val Gly Val Ser Asp Ala Ile Ser Asn Gly Tyr
        995                 1000                1005

GAC TCA TGG GGA CCT CTC TTT GGG AGA CTT TTC TTC GCT CTT TGG GTC      3310
Asp Ser Trp Gly Pro Leu Phe Gly Arg Leu Phe Phe Ala Leu Trp Val
    1010                1015                1020

ATT GTT CAT TTA TAC CCA TTC CTC AAG GGA ATG CTT GGG AAG CAA GAC      3358
Ile Val His Leu Tyr Pro Phe Leu Lys Gly Met Leu Gly Lys Gln Asp
1025                1030                1035                1040

AAA ATG CCT ACG ATT ATT GTG GTC TGG TCT ATT CTT CTA GCT TCG ATC      3406
Lys Met Pro Thr Ile Ile Val Val Trp Ser Ile Leu Leu Ala Ser Ile
                1045                1050                1055

TTG ACA CTC TTG TGG GTC AGA ATT AAC CCG TTT GTG GCT AAA GGG GGA      3454
Leu Thr Leu Leu Trp Val Arg Ile Asn Pro Phe Val Ala Lys Gly Gly
            1060                1065                1070
```

CCA GTG TTG GAG ATC TGT GGT CTG AAT TGT GGA AAC TAAGATCCTC 3500
Pro Val Leu Glu Ile Cys Gly Leu Asn Cys Gly Asn
1075 1080

AGTGAAAGAA GAGCAAAGGA GTTTGTGTTG GAGCTTTGGA AGCAAATGTG TTGATGATGA 3560

TGCAAGTGTG TTTGTAGACA AAGATGTGCA GTTTTTACTT TTTACGACTT GTTAAACCTT 3620

TTTTGTTACC CCTAAATTAA TTCTTTTGTT ATCATGGTTA TACTAATAGA ATTGTTTGTT 3680

TTTCTTTTTT ACATGTACTT TTAGTTATTC CGTAGTTATT GTATAATACT GATAACGATC 3740

ATATATACAC ACTTTGTTAA CAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAGCGGCCG 3800

CTCGAATTGT CGACGCGGCC GCGAATTC 3828

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1084 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Asn Thr Gly Gly Arg Leu Ile Ala Gly Ser His Asn Arg Asn Glu
1 5 10 15

Phe Val Leu Ile Asn Ala Asp Glu Ser Ala Arg Ile Arg Ser Val Gln
20 25 30

Glu Leu Ser Gly Gln Thr Cys Gln Ile Cys Gly Asp Glu Ile Glu Leu
35 40 45

```
Thr Val Ser Ser Glu Leu Phe Val Ala Cys Asn Glu Cys Ala Phe Pro
     50                  55                  60

Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Arg Glu Gly Asn Gln Ala
 65                  70                  75                  80

Cys Pro Gln Cys Lys Thr Arg Tyr Lys Arg Ile Lys Gly Ser Pro Arg
                 85                  90                  95

Val Asp Gly Asp Asp Glu Glu Glu Glu Asp Ile Asp Asp Leu Glu Tyr
            100                 105                 110

Glu Phe Asp His Gly Met Asp Pro Glu His Ala Ala Glu Ala Ala Leu
        115                 120                 125

Ser Ser Arg Leu Asn Thr Gly Arg Gly Gly Leu Asp Ser Ala Pro Pro
    130                 135                 140

Gly Ser Gln Ile Pro Leu Leu Thr Tyr Cys Asp Glu Asp Ala Asp Met
145                 150                 155                 160

Tyr Ser Asp Arg His Ala Leu Ile Val Pro Pro Ser Thr Gly Tyr Gly
                165                 170                 175

Asn Arg Val Tyr Pro Ala Pro Phe Thr Asp Ser Ser Ala Pro Pro Gln
            180                 185                 190

Ala Arg Ser Met Val Pro Gln Lys Asp Ile Ala Glu Tyr Gly Tyr Gly
        195                 200                 205

Ser Val Ala Trp Lys Asp Arg Met Glu Val Trp Lys Arg Arg Gln Gly
    210                 215                 220

Glu Lys Leu Gln Val Ile Lys His Glu Gly Gly Asn Asn Gly Arg Gly
225                 230                 235                 240

Ser Asn Asp Asp Asp Glu Leu Asp Asp Pro Asp Met Pro Met Met Asp
                245                 250                 255
```

Glu Gly Arg Gln Pro Leu Ser Arg Lys Leu Pro Ile Arg Ser Ser Arg
260 265 270

Ile Asn Pro Tyr Arg Met Leu Ile Leu Cys Arg Leu Ala Ile Leu Gly
275 280 285

Leu Phe Phe His Tyr Arg Ile Leu His Pro Val Asn Asp Ala Tyr Gly
290 295 300

Leu Trp Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Val Ser Trp
305 310 315 320

Ile Leu Asp Gln Phe Pro Lys Trp Tyr Pro Ile Glu Arg Glu Thr Tyr
325 330 335

Leu Asp Arg Leu Ser Leu Arg Tyr Glu Lys Glu Gly Lys Pro Ser Gly
340 345 350

Leu Ala Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu
355 360 365

Pro Pro Leu Ile Thr Ala Asn Thr Val Leu Ser Ile Leu Ala Val Asp
370 375 380

Tyr Pro Val Asp Lys Val Ala Cys Tyr Val Ser Asn Asn Gly Ala Ala
385 390 395 400

Met Leu Thr Phe Glu Ala Leu Ser Asp Thr Ala Asp Phe Ala Thr Lys
405 410 415

Trp Val Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro Glu
420 425 430

Trp Tyr Phe Ser Gln Lys Met Asp Tyr Leu Lys Asn Lys Val His Pro
435 440 445

Ala Phe Val Arg Glu Arg Arg Ala Met Lys Arg Asp Tyr Glu Glu Phe
450 455 460

```
Lys Val Lys Ile Asn Ala Leu Val Ala Thr Ala Gln Lys Val Pro Glu
465                 470                 475                 480

Glu Arg Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Val
                485                 490                 495

Arg Asp His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly Val
            500                 505                 510

Arg Asp Thr Asp Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg
        515                 520                 525

Glu Lys Arg Pro Gly Phe Asp His His Lys Lys Ala Gly Ala Met Asn
    530                 535                 540

Ser Leu Ile Arg Val Ser Ala Val Leu Ser Asn Ala Pro Tyr Leu Leu
545                 550                 555                 560

Asn Val Asp Cys Asp His Tyr Ile Asn Asn Ser Lys Ala Ile Arg Glu
                565                 570                 575

Ser Met Cys Phe Met Met Asp Pro Gln Ser Gly Lys Lys Val Cys Tyr
            580                 585                 590

Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Arg His Asp Arg Tyr
        595                 600                 605

Ser Asn Arg Asn Val Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp
    610                 615                 620

Gly Ile Gln Gly Pro Ile Tyr Val Gly Thr Gly Cys Val Phe Arg Lys
625                 630                 635                 640

Gln Ala Leu Tyr Gly Phe Asp Ala Pro Lys Lys Lys Lys Pro Pro Gly
                645                 650                 655

Lys Thr Cys Asn Cys Trp Pro Lys Trp Cys Cys Leu Cys Cys Gly Leu
            660                 665                 670
```

Arg Lys Lys Ser Lys Thr Lys Ala Thr Asp Lys Lys Thr Asn Thr Lys
675 680 685

Glu Thr Ser Lys Gln Ile His Ala Leu Glu Asn Val Asp Glu Gly Val
690 695 700

Ile Val Pro Val Ser Asn Val Glu Lys Arg Ser Glu Ala Thr Gln Leu
705 710 715 720

Lys Leu Glu Lys Lys Phe Gly Gln Ser Pro Val Phe Val Ala Ser Ala
725 730 735

Val Leu Gln Asn Gly Gly Val Pro Arg Asn Ala Ser Pro Ala Cys Leu
740 745 750

Leu Arg Glu Ala Ile Gln Val Ile Ser Cys Gly Tyr Gln Asp Lys Thr
755 760 765

Glu Trp Gly Lys Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp
770 775 780

Ile Leu Thr Gly Phe Lys Met His Cys His Gly Trp Arg Ser Val Tyr
785 790 795 800

Cys Met Pro Lys Arg Ala Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu
805 810 815

Ser Asp Arg Leu His Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu
820 825 830

Ile Phe Leu Ser Arg His Cys Pro Ile Trp Tyr Gly Tyr Gly Gly Gly
835 840 845

Leu Lys Trp Leu Glu Arg Phe Ser Tyr Ile Asn Ser Val Val Tyr Pro
850 855 860

Trp Thr Ser Leu Pro Leu Ile Val Tyr Cys Ser Leu Pro Ala Val Cys
865 870 875 880

Leu Leu Thr Gly Lys Phe Ile Val Pro Glu Ile Ser Asn Tyr Ala Gly
885 890 895

Ile Leu Phe Met Leu Met Phe Ile Ser Ile Ala Val Thr Gly Ile Leu
900 905 910

Glu Met Gln Trp Gly Gly Val Gly Ile Asp Asp Trp Trp Arg Asn Glu
915 920 925

Gln Phe Trp Val Ile Gly Gly Ala Ser Ser His Leu Phe Ala Leu Phe
930 935 940

Gln Gly Leu Leu Lys Val Leu Ala Gly Val Asn Thr Asn Phe Thr Val
945 950 955 960

Thr Ser Lys Ala Ala Asp Asp Gly Ala Phe Ser Glu Leu Tyr Ile Phe
965 970 975

Lys Trp Thr Thr Leu Leu Ile Pro Pro Thr Thr Leu Leu Ile Ile Asn
980 985 990

Ile Ile Gly Val Ile Val Gly Val Ser Asp Ala Ile Ser Asn Gly Tyr
995 1000 1005

Asp Ser Trp Gly Pro Leu Phe Gly Arg Leu Phe Phe Ala Leu Trp Val
1010 1015 1020

Ile Val His Leu Tyr Pro Phe Leu Lys Gly Met Leu Gly Lys Gln Asp
1025 1030 1035 1040

Lys Met Pro Thr Ile Ile Val Val Trp Ser Ile Leu Leu Ala Ser Ile
1045 1050 1055

Leu Thr Leu Leu Trp Val Arg Ile Asn Pro Phe Val Ala Lys Gly Gly
1060 1065 1070

Pro Val Leu Glu Ile Cys Gly Leu Asn Cys Gly Asn
1075 1080

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3614 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia

(vii) IMMEDIATE SOURCE:
(B) CLONE: Ath-B

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 217..3411

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GAATTCGCGG CCGCGTCGAC TACGGCTGCG AGAAGACGAC AGAAGGGGAT CCCAAGATTC 60

TCCTCTTCGT CTTCCTTATA AACTATCTCT CTGTAGAGAA GAAAGCTTGG ATCCAGATTG 120

AGAGAGATTC AGAGAGCCAC ATCACCACAC TCCATCTTCA GATCTCATGA TTTGAACTAT 180

TCCGACGTTT CGGTGTTGGA AGCAACTAAG TGACAA ATG GAA TCC GAA GGA GAA 234
Met Glu Ser Glu Gly Glu
1 5

ACC GCG GGA AAG CCG ATG AAG AAC ATT GTT CCG CAG ACT TGC CAG ATC 282
Thr Ala Gly Lys Pro Met Lys Asn Ile Val Pro Gln Thr Cys Gln Ile
10 15 20

```
TGT AGT GAC AAT GTT GGC AAG ACT GTT GAT GGA GAT CGT TTT GTG GCT      330
Cys Ser Asp Asn Val Gly Lys Thr Val Asp Gly Asp Arg Phe Val Ala
         25                  30                  35

TGT GAT ATT TGT TCA TTC CCA GTT TGT CGG CCT TGC TAC GAG TAT GAG      378
Cys Asp Ile Cys Ser Phe Pro Val Cys Arg Pro Cys Tyr Glu Tyr Glu
     40                  45                  50

AGG AAA GAT GGG AAT CAA TCT TGT CCT CAG TGC AAA ACC AGA TAC AAG      426
Arg Lys Asp Gly Asn Gln Ser Cys Pro Gln Cys Lys Thr Arg Tyr Lys
 55                  60                  65                  70

AGG CTC AAA GGT AGT CCT GCT ATT CCT GGT GAT AAA GAC GAG GAT GGC      474
Arg Leu Lys Gly Ser Pro Ala Ile Pro Gly Asp Lys Asp Glu Asp Gly
                 75                  80                  85

TTA GCT GAT GAA GGT ACT GTT GAG TTC AAC TAC CCT CAG AAG GAG AAA      522
Leu Ala Asp Glu Gly Thr Val Glu Phe Asn Tyr Pro Gln Lys Glu Lys
             90                  95                 100

ATT TCA GAG CGG ATG CTT GGT TGG CAT CTT ACT CGT GGG AAG GGA GAG      570
Ile Ser Glu Arg Met Leu Gly Trp His Leu Thr Arg Gly Lys Gly Glu
        105                 110                 115

GAA ATG GGG GAA CCC CAG TAT GAT AAA GAG GTC TCT CAC AAT CAT CTT      618
Glu Met Gly Glu Pro Gln Tyr Asp Lys Glu Val Ser His Asn His Leu
    120                 125                 130

CCT CGT CTC ACG AGC AGA CAA GAT ACT TCA GGA GAG TTT TCT GCT GCC      666
Pro Arg Leu Thr Ser Arg Gln Asp Thr Ser Gly Glu Phe Ser Ala Ala
135                 140                 145                 150

TCA CCT GAA CGC CTC TCT GTA TCT TCT ACT ATC GCT GGG GGA AAG CGC      714
Ser Pro Glu Arg Leu Ser Val Ser Ser Thr Ile Ala Gly Gly Lys Arg
                155                 160                 165

CTT CCC TAT TCA TCA GAT GTC AAT CAA TCA CCA AAT AGA AGG ATT GTG      762
Leu Pro Tyr Ser Ser Asp Val Asn Gln Ser Pro Asn Arg Arg Ile Val
            170                 175                 180
```

```
GAT CCT GTT GGA CTC GGG AAT GTA GCT TGG AAG GAG AGA GTT GAT GGC      810
Asp Pro Val Gly Leu Gly Asn Val Ala Trp Lys Glu Arg Val Asp Gly
        185                 190                 195

TGG AAA ATG AAG CAA GAG AAG AAT ACT GGT CCT GTC AGC ACG CAG GCT      858
Trp Lys Met Lys Gln Glu Lys Asn Thr Gly Pro Val Ser Thr Gln Ala
    200                 205                 210

GCT TCT GAA AGA GGT GGA GTA GAT ATT GAT GCC AGC ACA GAT ATC CTA      906
Ala Ser Glu Arg Gly Gly Val Asp Ile Asp Ala Ser Thr Asp Ile Leu
215                 220                 225                 230

GCA GAT GAG GCT CTG CTG AAT GAC GAA GCG AGG CAG CTT CTG TCA AGG      954
Ala Asp Glu Ala Leu Leu Asn Asp Glu Ala Arg Gln Leu Leu Ser Arg
                235                 240                 245

AAA GTT TCA ATT CCT TCA TCA CGG ATC AAT CCT TAC AGA ATG GTT ATT     1002
Lys Val Ser Ile Pro Ser Ser Arg Ile Asn Pro Tyr Arg Met Val Ile
            250                 255                 260

ATG CTG CGG CTT GTT ATC CTT TGT CTC TTC TTG CAT TAC CGT ATA ACA     1050
Met Leu Arg Leu Val Ile Leu Cys Leu Phe Leu His Tyr Arg Ile Thr
        265                 270                 275

AAC CCA GTG CCA AAT GCC TTT GCT CTA TGG CTG GTC TCT GTG ATA TGT     1098
Asn Pro Val Pro Asn Ala Phe Ala Leu Trp Leu Val Ser Val Ile Cys
    280                 285                 290

GAG ATC TGG TTT GCC TTA TCC TGG ATT TTG GAT CAG TTT CCC AAG TGG     1146
Glu Ile Trp Phe Ala Leu Ser Trp Ile Leu Asp Gln Phe Pro Lys Trp
295                 300                 305                 310

TTT CCT GTG AAC CGT GAA ACC TAC CTC GAC AGG CTT GCT TTA AGA TAT     1194
Phe Pro Val Asn Arg Glu Thr Tyr Leu Asp Arg Leu Ala Leu Arg Tyr
                315                 320                 325

GAT CGT GAA GGT GAG CCA TCA CAG TTA GCT GCT GTT GAC ATT TTC GTG     1242
Asp Arg Glu Gly Glu Pro Ser Gln Leu Ala Ala Val Asp Ile Phe Val
            330                 335                 340
```

```
AGT ACT GTT GAC CCC TTG AAG GAG CCA CCC CTT GTG ACA GCC AAC ACA      1290
Ser Thr Val Asp Pro Leu Lys Glu Pro Pro Leu Val Thr Ala Asn Thr
        345                 350                 355

GTG CTC TCT ATT CTG GCT GTT GAC TAC CCA GTT GAC AAG GTG TCC TGT      1338
Val Leu Ser Ile Leu Ala Val Asp Tyr Pro Val Asp Lys Val Ser Cys
    360                 365                 370

TAT GTT TCT GAT GAT GGT GCT GCT ATG TTA TCA TTT GAA TCA CTT GCA      1386
Tyr Val Ser Asp Asp Gly Ala Ala Met Leu Ser Phe Glu Ser Leu Ala
375                 380                 385                 390

GAA ACA TCA GAG TTT GCT CGT AAA TGG GTA CCA TTT TGC AAG AAA TAT      1434
Glu Thr Ser Glu Phe Ala Arg Lys Trp Val Pro Phe Cys Lys Lys Tyr
                395                 400                 405

AGC ATA GAG CCT CGT GCA CCA GAA TGG TAC TTT GCT GCG AAA ATA GAT      1482
Ser Ile Glu Pro Arg Ala Pro Glu Trp Tyr Phe Ala Ala Lys Ile Asp
            410                 415                 420

TAC TTG AAG GAT AAA GTT CAG ACA TCA TTT GTC AAA GAT CGT AGA GCT      1530
Tyr Leu Lys Asp Lys Val Gln Thr Ser Phe Val Lys Asp Arg Arg Ala
        425                 430                 435

ATG AAG AGG GAA TAT GAG GAA TTT AAA ATC CGA ATC AAT GCA CTT GTT      1578
Met Lys Arg Glu Tyr Glu Glu Phe Lys Ile Arg Ile Asn Ala Leu Val
    440                 445                 450

TCC AAA GCC CTA AAA TGT CCT GAA GAA GGG TGG GTT ATG CAA GAT GGC      1626
Ser Lys Ala Leu Lys Cys Pro Glu Glu Gly Trp Val Met Gln Asp Gly
455                 460                 465                 470

ACA CCG TGG CCT GGA AAT AAT ACA GGG GAC CAT CCA GGA ATG ATC CAG      1674
Thr Pro Trp Pro Gly Asn Asn Thr Gly Asp His Pro Gly Met Ile Gln
                475                 480                 485

GTC TTC TTA GGG CAA AAT GGT GGA CTT GAT GCA GAG GGC AAT GAG CTC      1722
Val Phe Leu Gly Gln Asn Gly Gly Leu Asp Ala Glu Gly Asn Glu Leu
            490                 495                 500
```

```
CCG CGT TTG GTA TAT GTT TCT CGA GAA AAG CGA CCA GGA TTC CAG CAC      1770
Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe Gln His
        505                 510                 515

CAC AAA AAG GCT GGT GCT ATG AAT GCA CTG GTG AGA GTT TCA GCA GTT      1818
His Lys Lys Ala Gly Ala Met Asn Ala Leu Val Arg Val Ser Ala Val
    520                 525                 530

CTT ACC AAT GGA CCT TTC ATC TTG AAT CTT GAT TGT GAT CAT TAC ATA      1866
Leu Thr Asn Gly Pro Phe Ile Leu Asn Leu Asp Cys Asp His Tyr Ile
535                 540                 545                 550

AAT AAC AGC AAA GCC TTA AGA GAA GCA ATG TGC TTC CTG ATG GAC CCA      1914
Asn Asn Ser Lys Ala Leu Arg Glu Ala Met Cys Phe Leu Met Asp Pro
                555                 560                 565

AAC CTC GGG AAG CAA GTT TGT TAT GTT CAG TTC CCA CAA AGA TTT GAT      1962
Asn Leu Gly Lys Gln Val Cys Tyr Val Gln Phe Pro Gln Arg Phe Asp
            570                 575                 580

GGT ATC GAT AAG AAC GAT AGA TAT GCT AAT CGT AAT ACC GTG TTC TTT      2010
Gly Ile Asp Lys Asn Asp Arg Tyr Ala Asn Arg Asn Thr Val Phe Phe
        585                 590                 595

GAT ATT AAC TTG AGA GGT TTA GAT GGG ATT CAA GGA CCT GTA TAT GTC      2058
Asp Ile Asn Leu Arg Gly Leu Asp Gly Ile Gln Gly Pro Val Tyr Val
    600                 605                 610

GGA ACT GGA TGT GTT TTC AAC AGA ACA GCA TTA TAC GGT TAT GAA CCT      2106
Gly Thr Gly Cys Val Phe Asn Arg Thr Ala Leu Tyr Gly Tyr Glu Pro
615                 620                 625                 630

CCA ATA AAA GTA AAA CAC AAG AAG CCA AGT CTT TTA TCT AAG CTC TGT      2154
Pro Ile Lys Val Lys His Lys Lys Pro Ser Leu Leu Ser Lys Leu Cys
                635                 640                 645

GGT GGA TCA AGA AAG AAG AAT TCC AAA GCT AAG AAA GAG TCG GAC AAA      2202
Gly Gly Ser Arg Lys Lys Asn Ser Lys Ala Lys Lys Glu Ser Asp Lys
            650                 655                 660
```

```
AAG AAA TCA GGC AGG CAT ACT GAC TCA ACT GTT CCT GTA TTC AAC CTC      2250
Lys Lys Ser Gly Arg His Thr Asp Ser Thr Val Pro Val Phe Asn Leu
        665                 670                 675

GAT GAC ATA GAA GAG GGA GTT GAA GGT GCT GGT TTT GAT GAT GAA AAG      2298
Asp Asp Ile Glu Glu Gly Val Glu Gly Ala Gly Phe Asp Asp Glu Lys
    680                 685                 690

GCG CTC TTA ATG TCG CAA ATG AGC CTG GAG AAG CGA TTT GGA CAG TCT      2346
Ala Leu Leu Met Ser Gln Met Ser Leu Glu Lys Arg Phe Gly Gln Ser
695                 700                 705                 710

GCT GTT TTT GTT GCT TCT ACC CTA ATG GAA AAT GGT GGT GTT CCT CCT      2394
Ala Val Phe Val Ala Ser Thr Leu Met Glu Asn Gly Gly Val Pro Pro
                715                 720                 725

TCA GCA ACT CCA GAA AAC TTT CTC AAA GAG GCT ATC CAT GTC ATT AGT      2442
Ser Ala Thr Pro Glu Asn Phe Leu Lys Glu Ala Ile His Val Ile Ser
            730                 735                 740

TGT GGT TAT GAG GAT AAG TCA GAT TGG GGA ATG GAG ATT GGA TGG ATC      2490
Cys Gly Tyr Glu Asp Lys Ser Asp Trp Gly Met Glu Ile Gly Trp Ile
        745                 750                 755

TAT GGT TCT GTG ACA GAA GAT ATT CTG ACT GGG TTC AAA ATG CAT GCC      2538
Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly Phe Lys Met His Ala
    760                 765                 770

CGT GGA TGG CGA TCC ATT TAC TGC ATG CCT AAG CTT CCA GCT TTC AAG      2586
Arg Gly Trp Arg Ser Ile Tyr Cys Met Pro Lys Leu Pro Ala Phe Lys
775                 780                 785                 790

GGT TCT GCT CCT ATC AAT CTT TCA GAT CGT CTG AAC CAA GTG CTG AGG      2634
Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu Asn Gln Val Leu Arg
                795                 800                 805

TGG GCT TTA GGT TCA GTT GAG ATT CTC TTC AGT CGG CAT TGT CCT ATA      2682
Trp Ala Leu Gly Ser Val Glu Ile Leu Phe Ser Arg His Cys Pro Ile
            810                 815                 820
```

```
TGG TAT GGT TAC AAT GGG AGG CTA AAA TTT CTT GAG AGG TTT GCG TAT      2730
Trp Tyr Gly Tyr Asn Gly Arg Leu Lys Phe Leu Glu Arg Phe Ala Tyr
        825                 830                 835

GTG AAC ACC ACC ATC TAC CCT ATC ACC TCC ATT CCT CTT CTC ATG TAT      2778
Val Asn Thr Thr Ile Tyr Pro Ile Thr Ser Ile Pro Leu Leu Met Tyr
    840                 845                 850

TGT ACA TTG CTA GCC GTT TGT CTC TTC ACC AAC CAG TTT ATT ATT CCT      2826
Cys Thr Leu Leu Ala Val Cys Leu Phe Thr Asn Gln Phe Ile Ile Pro
855                 860                 865                 870

CAG ATT AGT AAC ATT GCA AGT ATA TGG TTT CTG TCT CTC TTT CTC TCC      2874
Gln Ile Ser Asn Ile Ala Ser Ile Trp Phe Leu Ser Leu Phe Leu Ser
                875                 880                 885

ATT TTC GCC ACG GGT ATA CTA GAA ATG AGG TGG AGT GGC GTA GGC ATA      2922
Ile Phe Ala Thr Gly Ile Leu Glu Met Arg Trp Ser Gly Val Gly Ile
            890                 895                 900

GAC GAA TGG TGG AGA AAC GAG CAG TTT TGG GTC ATT GGT GGA GTA TCC      2970
Asp Glu Trp Trp Arg Asn Glu Gln Phe Trp Val Ile Gly Gly Val Ser
        905                 910                 915

GCT CAT TTA TTC GCT GTG TTT CAA GGT ATC CTC AAA GTC CTT GCC GGT      3018
Ala His Leu Phe Ala Val Phe Gln Gly Ile Leu Lys Val Leu Ala Gly
    920                 925                 930

ATT GAC ACA AAC TTC ACA GTT ACC TCA AAA GCT TCA GAT GAA GAC GGA      3066
Ile Asp Thr Asn Phe Thr Val Thr Ser Lys Ala Ser Asp Glu Asp Gly
935                 940                 945                 950

GAC TTT GCT GAG CTC TAC TTG TTC AAA TGG ACA ACA CTT CTG ATT CCG      3114
Asp Phe Ala Glu Leu Tyr Leu Phe Lys Trp Thr Thr Leu Leu Ile Pro
                955                 960                 965

CCA ACG ACG CTG CTC ATT GTA AAC TTA GTG GGA GTT GTT GCA GGA GTC      3162
Pro Thr Thr Leu Leu Ile Val Asn Leu Val Gly Val Val Ala Gly Val
            970                 975                 980
```

```
TCT TAT GCT ATC AAC AGT GGA TAC CAA TCA TGG GGA CCA CTC TTT GGT      3210
Ser Tyr Ala Ile Asn Ser Gly Tyr Gln Ser Trp Gly Pro Leu Phe Gly
        985                 990                 995

AAG TTG TTC TTT GCC TTC TGG GTG ATT GTT CAC TTG TAC CCT TTC CTC      3258
Lys Leu Phe Phe Ala Phe Trp Val Ile Val His Leu Tyr Pro Phe Leu
    1000                1005                1010

AAG GGT TTG ATG GGT CGA CAG AAC CGG ACT CCT ACC ATT GTT GTG GTC      3306
Lys Gly Leu Met Gly Arg Gln Asn Arg Thr Pro Thr Ile Val Val Val
1015                1020                1025                1030

TGG TCT GTT CTC TTG GCT TCT ATC TTC TCG TTG TTG TGG GTT AGG ATT      3354
Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu Leu Trp Val Arg Ile
                1035                1040                1045

GAT CCC TTC ACT AGC CGA GTC ACT GGC CCG GAC ATT CTG GAA TGT GGA      3402
Asp Pro Phe Thr Ser Arg Val Thr Gly Pro Asp Ile Leu Glu Cys Gly
            1050                1055                1060

ATC AAC TGT TGAGAAGCGA GCAAATATTT ACCTGTTTTG AGGGTTAAAA             3451
Ile Asn Cys
        1065

AAAACACAGA ATTTAAATTA TTTTTCATTG TTTTATTTGT TCACTTTTTT ACTTTTGTTG    3511

TGTGTATCTG TCTGTTCGTT CTTCTGTCTT GGTGTCATAA ATTTATGTGT AGAATATATC    3571

TTACTCTAGT TACTTTGGAA AGTTATAATT AAAGTGAAAG CCA                      3614
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1065 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Glu Ser Glu Gly Glu Thr Ala Gly Lys Pro Met Lys Asn Ile Val
 1               5                  10                  15

Pro Gln Thr Cys Gln Ile Cys Ser Asp Asn Val Gly Lys Thr Val Asp
            20                  25                  30

Gly Asp Arg Phe Val Ala Cys Asp Ile Cys Ser Phe Pro Val Cys Arg
        35                  40                  45

Pro Cys Tyr Glu Tyr Glu Arg Lys Asp Gly Asn Gln Ser Cys Pro Gln
    50                  55                  60

Cys Lys Thr Arg Tyr Lys Arg Leu Lys Gly Ser Pro Ala Ile Pro Gly
65                  70                  75                  80

Asp Lys Asp Glu Asp Gly Leu Ala Asp Glu Gly Thr Val Glu Phe Asn
                85                  90                  95

Tyr Pro Gln Lys Glu Lys Ile Ser Glu Arg Met Leu Gly Trp His Leu
            100                 105                 110

Thr Arg Gly Lys Gly Glu Glu Met Gly Glu Pro Gln Tyr Asp Lys Glu
        115                 120                 125

Val Ser His Asn His Leu Pro Arg Leu Thr Ser Arg Gln Asp Thr Ser
    130                 135                 140

Gly Glu Phe Ser Ala Ala Ser Pro Glu Arg Leu Ser Val Ser Ser Thr
145                 150                 155                 160

Ile Ala Gly Gly Lys Arg Leu Pro Tyr Ser Ser Asp Val Asn Gln Ser
                165                 170                 175

Pro Asn Arg Arg Ile Val Asp Pro Val Gly Leu Gly Asn Val Ala Trp
            180                 185                 190
```

Lys Glu Arg Val Asp Gly Trp Lys Met Lys Gln Glu Lys Asn Thr Gly
195 200 205

Pro Val Ser Thr Gln Ala Ala Ser Glu Arg Gly Gly Val Asp Ile Asp
210 215 220

Ala Ser Thr Asp Ile Leu Ala Asp Glu Ala Leu Leu Asn Asp Glu Ala
225 230 235 240

Arg Gln Leu Leu Ser Arg Lys Val Ser Ile Pro Ser Ser Arg Ile Asn
245 250 255

Pro Tyr Arg Met Val Ile Met Leu Arg Leu Val Ile Leu Cys Leu Phe
260 265 270

Leu His Tyr Arg Ile Thr Asn Pro Val Pro Asn Ala Phe Ala Leu Trp
275 280 285

Leu Val Ser Val Ile Cys Glu Ile Trp Phe Ala Leu Ser Trp Ile Leu
290 295 300

Asp Gln Phe Pro Lys Trp Phe Pro Val Asn Arg Glu Thr Tyr Leu Asp
305 310 315 320

Arg Leu Ala Leu Arg Tyr Asp Arg Glu Gly Glu Pro Ser Gln Leu Ala
325 330 335

Ala Val Asp Ile Phe Val Ser Thr Val Asp Pro Leu Lys Glu Pro Pro
340 345 350

Leu Val Thr Ala Asn Thr Val Leu Ser Ile Leu Ala Val Asp Tyr Pro
355 360 365

Val Asp Lys Val Ser Cys Tyr Val Ser Asp Asp Gly Ala Ala Met Leu
370 375 380

Ser Phe Glu Ser Leu Ala Glu Thr Ser Glu Phe Ala Arg Lys Trp Val
385 390 395 400

```
Pro Phe Cys Lys Lys Tyr Ser Ile Glu Pro Arg Ala Pro Glu Trp Tyr
                405                 410                 415

Phe Ala Ala Lys Ile Asp Tyr Leu Lys Asp Lys Val Gln Thr Ser Phe
            420                 425                 430

Val Lys Asp Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Ile
        435                 440                 445

Arg Ile Asn Ala Leu Val Ser Lys Ala Leu Lys Cys Pro Glu Glu Gly
    450                 455                 460

Trp Val Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Gly Asp
465                 470                 475                 480

His Pro Gly Met Ile Gln Val Phe Leu Gly Gln Asn Gly Gly Leu Asp
                485                 490                 495

Ala Glu Gly Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg Glu Lys
            500                 505                 510

Arg Pro Gly Phe Gln His His Lys Lys Ala Gly Ala Met Asn Ala Leu
        515                 520                 525

Val Arg Val Ser Ala Val Leu Thr Asn Gly Pro Phe Ile Leu Asn Leu
    530                 535                 540

Asp Cys Asp His Tyr Ile Asn Asn Ser Lys Ala Leu Arg Glu Ala Met
545                 550                 555                 560

Cys Phe Leu Met Asp Pro Asn Leu Gly Lys Gln Val Cys Tyr Val Gln
                565                 570                 575

Phe Pro Gln Arg Phe Asp Gly Ile Asp Lys Asn Asp Arg Tyr Ala Asn
            580                 585                 590

Arg Asn Thr Val Phe Phe Asp Ile Asn Leu Arg Gly Leu Asp Gly Ile
        595                 600                 605
```

```
Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Val Phe Asn Arg Thr Ala
    610                 615                 620

Leu Tyr Gly Tyr Glu Pro Pro Ile Lys Val Lys His Lys Lys Pro Ser
625                 630                 635                 640

Leu Leu Ser Lys Leu Cys Gly Gly Ser Arg Lys Lys Asn Ser Lys Ala
                645                 650                 655

Lys Lys Glu Ser Asp Lys Lys Lys Ser Gly Arg His Thr Asp Ser Thr
            660                 665                 670

Val Pro Val Phe Asn Leu Asp Asp Ile Glu Glu Gly Val Glu Gly Ala
        675                 680                 685

Gly Phe Asp Asp Glu Lys Ala Leu Leu Met Ser Gln Met Ser Leu Glu
    690                 695                 700

Lys Arg Phe Gly Gln Ser Ala Val Phe Val Ala Ser Thr Leu Met Glu
705                 710                 715                 720

Asn Gly Gly Val Pro Pro Ser Ala Thr Pro Glu Asn Phe Leu Lys Glu
                725                 730                 735

Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Ser Asp Trp Gly
            740                 745                 750

Met Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr
        755                 760                 765

Gly Phe Lys Met His Ala Arg Gly Trp Arg Ser Ile Tyr Cys Met Pro
    770                 775                 780

Lys Leu Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg
785                 790                 795                 800

Leu Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Val Glu Ile Leu Phe
                805                 810                 815
```

Ser Arg His Cys Pro Ile Trp Tyr Gly Tyr Asn Gly Arg Leu Lys Phe
820 825 830

Leu Glu Arg Phe Ala Tyr Val Asn Thr Thr Ile Tyr Pro Ile Thr Ser
835 840 845

Ile Pro Leu Leu Met Tyr Cys Thr Leu Leu Ala Val Cys Leu Phe Thr
850 855 860

Asn Gln Phe Ile Ile Pro Gln Ile Ser Asn Ile Ala Ser Ile Trp Phe
865 870 875 880

Leu Ser Leu Phe Leu Ser Ile Phe Ala Thr Gly Ile Leu Glu Met Arg
885 890 895

Trp Ser Gly Val Gly Ile Asp Glu Trp Trp Arg Asn Glu Gln Phe Trp
900 905 910

Val Ile Gly Gly Val Ser Ala His Leu Phe Ala Val Phe Gln Gly Ile
915 920 925

Leu Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr Ser Lys
930 935 940

Ala Ser Asp Glu Asp Gly Asp Phe Ala Glu Leu Tyr Leu Phe Lys Trp
945 950 955 960

Thr Thr Leu Leu Ile Pro Pro Thr Thr Leu Leu Ile Val Asn Leu Val
965 970 975

Gly Val Val Ala Gly Val Ser Tyr Ala Ile Asn Ser Gly Tyr Gln Ser
980 985 990

Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe Ala Phe Trp Val Ile Val
995 1000 1005

His Leu Tyr Pro Phe Leu Lys Gly Leu Met Gly Arg Gln Asn Arg Thr
1010 1015 1020

Pro Thr Ile Val Val Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser
1025 1030 1035 1040

Leu Leu Trp Val Arg Ile Asp Pro Phe Thr Ser Arg Val Thr Gly Pro
1045 1050 1055

Asp Ile Leu Glu Cys Gly Ile Asn Cys
1060 1065

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3673 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Arabidopsis thaliana
(B) STRAIN: Columbia
(C) INDIVIDUAL ISOLATE: *rsw1* mutant

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 71..3313

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GAATCGGCTA CGAATTTCCC AATTTTGAAT TTTGTGAATC TCTCTCTTTC TCTGTGTGTC 60

```
GGTGGCTGCG ATG GAG GCC AGT GCC GGC TTG GTT GCT GGA TCC TAC CGG        109
           Met Glu Ala Ser Ala Gly Leu Val Ala Gly Ser Tyr Arg
             1               5                  10

AGA AAC GAG CTC GTT CGG ATC CGA CAT GAA TCT GAT GGC GGG ACC AAA       157
Arg Asn Glu Leu Val Arg Ile Arg His Glu Ser Asp Gly Gly Thr Lys
     15                  20                  25

CCT TTG AAG AAT ATG AAT GGC CAG ATA TGT CAG ATC TGT GGT GAT GAT       205
Pro Leu Lys Asn Met Asn Gly Gln Ile Cys Gln Ile Cys Gly Asp Asp
 30                  35                  40                  45

GTT GGA CTC GCT GAA ACT GGA GAT GTC TTT GTC GCG TGT AAT GAA TGT       253
Val Gly Leu Ala Glu Thr Gly Asp Val Phe Val Ala Cys Asn Glu Cys
                 50                  55                  60

GCC TTC CCT GTG TGT CGG CCT TGC TAT GAG TAC GAG AGG AAA GAT GGA       301
Ala Phe Pro Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Lys Asp Gly
             65                  70                  75

ACT CAG TGT TGC CCT CAA TGC AAG ACT AGA TTC AGA CGA CAC AGG GGG       349
Thr Gln Cys Cys Pro Gln Cys Lys Thr Arg Phe Arg Arg His Arg Gly
         80                  85                  90

AGT CCT CGT GTT GAA GGA GAT GAA GAT GAG GAT GAT GTT GAT GAT ATC       397
Ser Pro Arg Val Glu Gly Asp Glu Asp Glu Asp Asp Val Asp Asp Ile
     95                  100                 105

GAG AAT GAG TTC AAT TAC GCC CAG GGA GCT AAC AAG GCG AGA CAC CAA       445
Glu Asn Glu Phe Asn Tyr Ala Gln Gly Ala Asn Lys Ala Arg His Gln
110                 115                 120                 125

CGC CAT GGC GAA GAG TTT TCT TCT TCC TCT AGA CAT GAA TCT CAA CCA       493
Arg His Gly Glu Glu Phe Ser Ser Ser Ser Arg His Glu Ser Gln Pro
                130                 135                 140

ATT CCT CTT CTC ACC CAT GGC CAT ACG GTT TCT GGA GAG ATT CGC ACG       541
Ile Pro Leu Leu Thr His Gly His Thr Val Ser Gly Glu Ile Arg Thr
            145                 150                 155
```

```
CCT GAT ACA CAA TCT GTG CGA ACT ACA TCA GGT CCT TTG GGT CCT TCT      589
Pro Asp Thr Gln Ser Val Arg Thr Thr Ser Gly Pro Leu Gly Pro Ser
        160                 165                 170

GAC AGG AAT GCT ATT TCA TCT CCA TAT ATT GAT CCA CGG CAA CCT GTC      637
Asp Arg Asn Ala Ile Ser Ser Pro Tyr Ile Asp Pro Arg Gln Pro Val
    175                 180                 185

CCT GTA AGA ATC GTG GAC CCG TCA AAA GAC TTG AAC TCT TAT GGG CTT      685
Pro Val Arg Ile Val Asp Pro Ser Lys Asp Leu Asn Ser Tyr Gly Leu
190                 195                 200                 205

GGT AAT GTT GAC TGG AAA GAA AGA GTT GAA GGC TGG AAG CTG AAG CAG      733
Gly Asn Val Asp Trp Lys Glu Arg Val Glu Gly Trp Lys Leu Lys Gln
                210                 215                 220

GAG AAA AAT ATG TTA CAG ATG ACT GGT AAA TAC CAT GAA GGG AAA GGA      781
Glu Lys Asn Met Leu Gln Met Thr Gly Lys Tyr His Glu Gly Lys Gly
            225                 230                 235

GGA GAA ATT GAA GGG ACT GGT TCC AAT GGC GAA GAA CTC CAA ATG GCT      829
Gly Glu Ile Glu Gly Thr Gly Ser Asn Gly Glu Glu Leu Gln Met Ala
        240                 245                 250

GAT GAT ACA CGT CTT CCT ATG AGT CGT GTG GTG CCT ATC CCA TCT TCT      877
Asp Asp Thr Arg Leu Pro Met Ser Arg Val Val Pro Ile Pro Ser Ser
    255                 260                 265

CGC CTA ACC CCT TAT CGG GTT GTG ATT ATT CTC CGG CTT ATC ATC TTG      925
Arg Leu Thr Pro Tyr Arg Val Val Ile Ile Leu Arg Leu Ile Ile Leu
270                 275                 280                 285

TGT TTC TTC TTG CAA TAT CGT ACA ACT CAC CCT GTG AAA AAT GCA TAT      973
Cys Phe Phe Leu Gln Tyr Arg Thr Thr His Pro Val Lys Asn Ala Tyr
                290                 295                 300

CCT TTG TGG TTG ACC TCG GTT ATC TGT GAG ATC TGG TTT GCA TTT TCT     1021
Pro Leu Trp Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Phe Ser
            305                 310                 315
```

```
TGG CTT CTT GAT CAG TTT CCC AAA TGG TAC CCC ATT AAC AGG GAG ACT      1069
Trp Leu Leu Asp Gln Phe Pro Lys Trp Tyr Pro Ile Asn Arg Glu Thr
        320                 325                 330

TAT CTT GAC CGT CTC GCT ATA AGA TAT GAT CGA GAC GGT GAA CCA TCA      1117
Tyr Leu Asp Arg Leu Ala Ile Arg Tyr Asp Arg Asp Gly Glu Pro Ser
    335                 340                 345

CAG CTC GTT CCT GTT GAT GTG TTT GTT AGT ACA GTG GAC CCA TTG AAA      1165
Gln Leu Val Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys
350                 355                 360                 365

GAG CCT CCC CTT GTT ACA GCA AAC ACA GTT CTC TCG ATT CTT TCT GTG      1213
Glu Pro Pro Leu Val Thr Ala Asn Thr Val Leu Ser Ile Leu Ser Val
                370                 375                 380

GAC TAC CCG GTA GAT AAA GTA GCC TGT TAT GTT TCA GAT GAT GGT TCA      1261
Asp Tyr Pro Val Asp Lys Val Ala Cys Tyr Val Ser Asp Asp Gly Ser
            385                 390                 395

GCT ATG CTT ACC TTT GAA TCC CTT TCT GAA ACC GCT GAG TTT GCA AAG      1309
Ala Met Leu Thr Phe Glu Ser Leu Ser Glu Thr Ala Glu Phe Ala Lys
        400                 405                 410

AAA TGG GTA CCA TTT TGC AAG AAA TTC AAC ATT GAA CCT AGG GCC CCT      1357
Lys Trp Val Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro
    415                 420                 425

GAA TTC TAT TTT GCC CAG AAG ATA GAT TAC TTG AAG GAC AAG ATC CAA      1405
Glu Phe Tyr Phe Ala Gln Lys Ile Asp Tyr Leu Lys Asp Lys Ile Gln
430                 435                 440                 445

CCG TCT TTT GTT AAA GAG CGA CGA GCT ATG AAG AGA GAG TAT GAA GAG      1453
Pro Ser Phe Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu
                450                 455                 460

TTT AAA GTG AGG ATA AAT GCT CTT GTT GCC AAA GCA CAG AAA ATC CCT      1501
Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Ile Pro
            465                 470                 475
```

GAA GAA GGC TGG ACA ATG CAG GAT GGT ACT CCC TGG CCT GGT AAC AAC 1549
Glu Glu Gly Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn
480 485 490

ACT AGA GAT CAT CCT GGA ATG ATA CAG GTG TTC TTA GGC CAT AGT GGG 1597
Thr Arg Asp His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly
495 500 505

GGT CTG GAT ACC GAT GGA AAT GAG CTG CCT AGA CTC ATC TAT GTT TCT 1645
Gly Leu Asp Thr Asp Gly Asn Glu Leu Pro Arg Leu Ile Tyr Val Ser
510 515 520 525

CGT GAA AAG CGG CCT GGA TTT CAA CAC CAC AAA AAG GCT GGA GCT ATG 1693
Arg Glu Lys Arg Pro Gly Phe Gln His His Lys Lys Ala Gly Ala Met
530 535 540

AAT GCA TTG ATC CGT GTA TCT GTT GTT CTT ACC AAT GGA GCA TAT CTT 1741
Asn Ala Leu Ile Arg Val Ser Val Val Leu Thr Asn Gly Ala Tyr Leu
545 550 555

TTG AAC GTG GAT TGT GAT CAT TAC TTT AAT AAC AGT AAG GCT ATT AAA 1789
Leu Asn Val Asp Cys Asp His Tyr Phe Asn Asn Ser Lys Ala Ile Lys
560 565 570

GAA GCT ATG TGT TTC ATG ATG GAC CCG GCT ATT GGA AAG AAG TGC TGC 1837
Glu Ala Met Cys Phe Met Met Asp Pro Ala Ile Gly Lys Lys Cys Cys
575 580 585

TAT GTC CAG TTC CCT CAA CGT TTT GAC GGT ATT GAT TTG CAC GAT CGA 1885
Tyr Val Gln Phe Pro Gln Arg Phe Asp Gly Ile Asp Leu His Asp Arg
590 595 600 605

TAT GCC AAC AGG AAT ATA GTC TTT TTC GAT ATT AAC ATG AAG GGG TTG 1933
Tyr Ala Asn Arg Asn Ile Val Phe Phe Asp Ile Asn Met Lys Gly Leu
610 615 620

GAT GGT ATC CAG GGT CCA GTA TAT GTG GGT ACT GGT TGT TGT TTT AAT 1981
Asp Gly Ile Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Cys Phe Asn
625 630 635

```
AGG CAG GCT CTA TAT GGG TAT GAT CCT GTT TTG ACG GAA GAA GAT TTA      2029
Arg Gln Ala Leu Tyr Gly Tyr Asp Pro Val Leu Thr Glu Glu Asp Leu
        640                 645                 650

GAA CCA AAT ATT ATT GTC AAG AGC TGT TGC GGG TCA AGG AAG AAA GGT      2077
Glu Pro Asn Ile Ile Val Lys Ser Cys Cys Gly Ser Arg Lys Lys Gly
    655                 660                 665

AAA AGT AGC AAG AAG TAT AAC TAC GAA AAG AGG AGA GGC ATC AAC AGA      2125
Lys Ser Ser Lys Lys Tyr Asn Tyr Glu Lys Arg Arg Gly Ile Asn Arg
670                 675                 680                 685

AGT GAC TCC AAT GCT CCA CTT TTC AAT ATG GAG GAC ATC GAT GAG GGT      2173
Ser Asp Ser Asn Ala Pro Leu Phe Asn Met Glu Asp Ile Asp Glu Gly
                690                 695                 700

TTT GAA GGT TAT GAT GAT GAG AGG TCT ATT CTA ATG TCC CAG AGG AGT      2221
Phe Glu Gly Tyr Asp Asp Glu Arg Ser Ile Leu Met Ser Gln Arg Ser
            705                 710                 715

GTA GAG AAG CGT TTT GGT CAG TCG CCG GTA TTT ATT GCG GCA ACC TTC      2269
Val Glu Lys Arg Phe Gly Gln Ser Pro Val Phe Ile Ala Ala Thr Phe
        720                 725                 730

ATG GAA CAA GGC GGC ATT CCA CCA ACA ACC AAT CCC GCT ACT CTT CTG      2317
Met Glu Gln Gly Gly Ile Pro Pro Thr Thr Asn Pro Ala Thr Leu Leu
    735                 740                 745

AAG GAG GCT ATT CAT GTT ATA AGC TGT GGT TAC GAA GAC AAG ACT GAA      2365
Lys Glu Ala Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu
750                 755                 760                 765

TGG GGC AAA GAG ATT GGT TGG ATC TAT GGT TCC GTG ACG GAA GAT ATT      2413
Trp Gly Lys Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile
                770                 775                 780

CTT ACT GGG TTC AAG ATG CAT GCC CGG GGT TGG ATA TCG ATC TAC TGC      2461
Leu Thr Gly Phe Lys Met His Ala Arg Gly Trp Ile Ser Ile Tyr Cys
            785                 790                 795
```

```
AAT CCT CCA CGC CCT GCG TTC AAG GGA TCT GCA CCA ATC AAT CTT TCT      2509
Asn Pro Pro Arg Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser
        800                 805                 810

GAT CGT TTG AAC CAA GTT CTT CGA TGG GCT TTG GGA TCT ATC GAG ATT      2557
Asp Arg Leu Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Ile Glu Ile
    815                 820                 825

CTT CTT AGC AGA CAT TGT CCT ATC TGG TAT GGT TAC CAT GGA AGG TTG      2605
Leu Leu Ser Arg His Cys Pro Ile Trp Tyr Gly Tyr His Gly Arg Leu
830                 835                 840                 845

AGA CTT TTG GAG AGG ATC GCT TAT ATC AAC ACC ATC GTC TAT CCT ATT      2653
Arg Leu Leu Glu Arg Ile Ala Tyr Ile Asn Thr Ile Val Tyr Pro Ile
                850                 855                 860

ACA TCC ATC CCT CTT ATT GCG TAT TGT ATT CTT CCC GCT TTT TGT CTC      2701
Thr Ser Ile Pro Leu Ile Ala Tyr Cys Ile Leu Pro Ala Phe Cys Leu
            865                 870                 875

ATC ACC GAC AGA TTC ATC ATA CCC GAG ATA AGC AAC TAC GCG AGT ATT      2749
Ile Thr Asp Arg Phe Ile Ile Pro Glu Ile Ser Asn Tyr Ala Ser Ile
        880                 885                 890

TGG TTC ATT CTA CTC TTC ATC TCA ATT GCT GTG ACT GGA ATC CTG GAG      2797
Trp Phe Ile Leu Leu Phe Ile Ser Ile Ala Val Thr Gly Ile Leu Glu
    895                 900                 905

CTG AGA TGG AGC GGT GTG AGC ATT GAG GAT TGG TGG AGG AAC GAG CAG      2845
Leu Arg Trp Ser Gly Val Ser Ile Glu Asp Trp Trp Arg Asn Glu Gln
910                 915                 920                 925

TTC TGG GTC ATT GGT GGC ACA TCC GCC CAT CTT TTT GCT GTC TTC CAA      2893
Phe Trp Val Ile Gly Gly Thr Ser Ala His Leu Phe Ala Val Phe Gln
                930                 935                 940

GGT CTA CTT AAG GTT CTT GCT GGT ATC GAC ACC AAC TTC ACC GTT ACA      2941
Gly Leu Leu Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr
            945                 950                 955
```

```
TCT AAA GCC ACA GAC GAA GAT GGG GAT TTT GCA GAA CTC TAC ATC TTC      2989
Ser Lys Ala Thr Asp Glu Asp Gly Asp Phe Ala Glu Leu Tyr Ile Phe
        960                 965                 970

AAA TGG ACA GCT CTT CTC ATT CCA CCA ACC ACC GTC CTA CTT GTG AAC      3037
Lys Trp Thr Ala Leu Leu Ile Pro Pro Thr Thr Val Leu Leu Val Asn
    975                 980                 985

CTC ATA GGC ATT GTG GCT GGT GTC TCT TAT GCT GTA AAC AGT GGC TAC      3085
Leu Ile Gly Ile Val Ala Gly Val Ser Tyr Ala Val Asn Ser Gly Tyr
990                 995                 1000                1005

CAG TCG TGG GGT CCG CTT TTC GGG AAG CTC TTC TTC GCC TTA TGG GTT      3133
Gln Ser Trp Gly Pro Leu Phe Gly Lys Leu Phe Phe Ala Leu Trp Val
                1010                1015                1020

ATT GCC CAT CTC TAC CCT TTC TTG AAA GGT CTG TTG GGA AGA CAA AAC      3181
Ile Ala His Leu Tyr Pro Phe Leu Lys Gly Leu Leu Gly Arg Gln Asn
            1025                1030                1035

CGA ACA CCA ACC ATC GTC ATT GTC TGG TCT GTT CTT CTC GCC TCC ATC      3229
Arg Thr Pro Thr Ile Val Ile Val Trp Ser Val Leu Leu Ala Ser Ile
        1040                1045                1050

TTC TCG TTG CTT TGG GTC AGG ATC AAT CCC TTT GTG GAC GCC AAT CCC      3277
Phe Ser Leu Leu Trp Val Arg Ile Asn Pro Phe Val Asp Ala Asn Pro
    1055                1060                1065

AAT GCC AAC AAC TTC AAT GGC AAA GGA GGT GTC TTT TAGACCCTAT          3323
Asn Ala Asn Asn Phe Asn Gly Lys Gly Gly Val Phe
1070                1075                1080

TTATATACTT GTGTGTGCAT ATATCAAAAA CGCGCAATGG GAATTCCAAA TCATCTAAAC    3383

CCATCAAACC CCAGTGAACC GGGCAGTTAA GGTGATTCCA TGTCCAAGAT TAGCTTTCTC    3443

CGAGTAGCCA GAGAAGGTGA AATTGTTCGT AACACTATTG TAATGATTTT CCAGTGGGGA    3503

AGAAGATGTG GACCCAAATG ATACATAGTC TACAAAAAGA ATTTGTTATT CTTTCTTATA    3563
```

TTTATTTTAT TTAAAGCTTG TTAGACTCAC ACTTATGTAA TGTTGGAACT TGTTGTCCTA 3623

AAAAGGGATT GGAGTTTTCT TTTTATCTAA GAATCTGAAG TTTATATGCT 3673

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1081 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Glu Ala Ser Ala Gly Leu Val Ala Gly Ser Tyr Arg Arg Asn Glu
1 5 10 15

Leu Val Arg Ile Arg His Glu Ser Asp Gly Gly Thr Lys Pro Leu Lys
20 25 30

Asn Met Asn Gly Gln Ile Cys Gln Ile Cys Gly Asp Asp Val Gly Leu
35 40 45

Ala Glu Thr Gly Asp Val Phe Val Ala Cys Asn Glu Cys Ala Phe Pro
50 55 60

Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Lys Asp Gly Thr Gln Cys
65 70 75 80

Cys Pro Gln Cys Lys Thr Arg Phe Arg Arg His Arg Gly Ser Pro Arg
85 90 95

Val Glu Gly Asp Glu Asp Glu Asp Asp Val Asp Asp Ile Glu Asn Glu
100 105 110

```
Phe Asn Tyr Ala Gln Gly Ala Asn Lys Ala Arg His Gln Arg His Gly
        115                 120                 125

Glu Glu Phe Ser Ser Ser Ser Arg His Glu Ser Gln Pro Ile Pro Leu
    130                 135                 140

Leu Thr His Gly His Thr Val Ser Gly Glu Ile Arg Thr Pro Asp Thr
145                 150                 155                 160

Gln Ser Val Arg Thr Thr Ser Gly Pro Leu Gly Pro Ser Asp Arg Asn
                165                 170                 175

Ala Ile Ser Ser Pro Tyr Ile Asp Pro Arg Gln Pro Val Pro Val Arg
            180                 185                 190

Ile Val Asp Pro Ser Lys Asp Leu Asn Ser Tyr Gly Leu Gly Asn Val
        195                 200                 205

Asp Trp Lys Glu Arg Val Glu Gly Trp Lys Leu Lys Gln Glu Lys Asn
    210                 215                 220

Met Leu Gln Met Thr Gly Lys Tyr His Glu Gly Lys Gly Gly Glu Ile
225                 230                 235                 240

Glu Gly Thr Gly Ser Asn Gly Glu Glu Leu Gln Met Ala Asp Asp Thr
                245                 250                 255

Arg Leu Pro Met Ser Arg Val Val Pro Ile Pro Ser Ser Arg Leu Thr
            260                 265                 270

Pro Tyr Arg Val Val Ile Ile Leu Arg Leu Ile Ile Leu Cys Phe Phe
        275                 280                 285

Leu Gln Tyr Arg Thr Thr His Pro Val Lys Asn Ala Tyr Pro Leu Trp
    290                 295                 300

Leu Thr Ser Val Ile Cys Glu Ile Trp Phe Ala Phe Ser Trp Leu Leu
305                 310                 315                 320
```

Asp Gln Phe Pro Lys Trp Tyr Pro Ile Asn Arg Glu Thr Tyr Leu Asp
325 330 335

Arg Leu Ala Ile Arg Tyr Asp Arg Asp Gly Glu Pro Ser Gln Leu Val
340 345 350

Pro Val Asp Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu Pro Pro
355 360 365

Leu Val Thr Ala Asn Thr Val Leu Ser Ile Leu Ser Val Asp Tyr Pro
370 375 380

Val Asp Lys Val Ala Cys Tyr Val Ser Asp Asp Gly Ser Ala Met Leu
385 390 395 400

Thr Phe Glu Ser Leu Ser Glu Thr Ala Glu Phe Ala Lys Lys Trp Val
405 410 415

Pro Phe Cys Lys Lys Phe Asn Ile Glu Pro Arg Ala Pro Glu Phe Tyr
420 425 430

Phe Ala Gln Lys Ile Asp Tyr Leu Lys Asp Lys Ile Gln Pro Ser Phe
435 440 445

Val Lys Glu Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val
450 455 460

Arg Ile Asn Ala Leu Val Ala Lys Ala Gln Lys Ile Pro Glu Glu Gly
465 470 475 480

Trp Thr Met Gln Asp Gly Thr Pro Trp Pro Gly Asn Asn Thr Arg Asp
485 490 495

His Pro Gly Met Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp
500 505 510

Thr Asp Gly Asn Glu Leu Pro Arg Leu Ile Tyr Val Ser Arg Glu Lys
515 520 525

Arg Pro Gly Phe Gln His His Lys Lys Ala Gly Ala Met Asn Ala Leu
530 535 540

Ile Arg Val Ser Val Val Leu Thr Asn Gly Ala Tyr Leu Leu Asn Val
545 550 555 560

Asp Cys Asp His Tyr Phe Asn Asn Ser Lys Ala Ile Lys Glu Ala Met
565 570 575

Cys Phe Met Met Asp Pro Ala Ile Gly Lys Lys Cys Cys Tyr Val Gln
580 585 590

Phe Pro Gln Arg Phe Asp Gly Ile Asp Leu His Asp Arg Tyr Ala Asn
595 600 605

Arg Asn Ile Val Phe Phe Asp Ile Asn Met Lys Gly Leu Asp Gly Ile
610 615 620

Gln Gly Pro Val Tyr Val Gly Thr Gly Cys Cys Phe Asn Arg Gln Ala
625 630 635 640

Leu Tyr Gly Tyr Asp Pro Val Leu Thr Glu Glu Asp Leu Glu Pro Asn
645 650 655

Ile Ile Val Lys Ser Cys Cys Gly Ser Arg Lys Lys Gly Lys Ser Ser
660 665 670

Lys Lys Tyr Asn Tyr Glu Lys Arg Arg Gly Ile Asn Arg Ser Asp Ser
675 680 685

Asn Ala Pro Leu Phe Asn Met Glu Asp Ile Asp Glu Gly Phe Glu Gly
690 695 700

Tyr Asp Asp Glu Arg Ser Ile Leu Met Ser Gln Arg Ser Val Glu Lys
705 710 715 720

Arg Phe Gly Gln Ser Pro Val Phe Ile Ala Ala Thr Phe Met Glu Gln
725 730 735

Gly Gly Ile Pro Pro Thr Thr Asn Pro Ala Thr Leu Leu Lys Glu Ala
740 745 750

Ile His Val Ile Ser Cys Gly Tyr Glu Asp Lys Thr Glu Trp Gly Lys
755 760 765

Glu Ile Gly Trp Ile Tyr Gly Ser Val Thr Glu Asp Ile Leu Thr Gly
770 775 780

Phe Lys Met His Ala Arg Gly Trp Ile Ser Ile Tyr Cys Asn Pro Pro
785 790 795 800

Arg Pro Ala Phe Lys Gly Ser Ala Pro Ile Asn Leu Ser Asp Arg Leu
805 810 815

Asn Gln Val Leu Arg Trp Ala Leu Gly Ser Ile Glu Ile Leu Leu Ser
820 825 830

Arg His Cys Pro Ile Trp Tyr Gly Tyr His Gly Arg Leu Arg Leu Leu
835 840 845

Glu Arg Ile Ala Tyr Ile Asn Thr Ile Val Tyr Pro Ile Thr Ser Ile
850 855 860

Pro Leu Ile Ala Tyr Cys Ile Leu Pro Ala Phe Cys Leu Ile Thr Asp
865 870 875 880

Arg Phe Ile Ile Pro Glu Ile Ser Asn Tyr Ala Ser Ile Trp Phe Ile
885 890 895

Leu Leu Phe Ile Ser Ile Ala Val Thr Gly Ile Leu Glu Leu Arg Trp
900 905 910

Ser Gly Val Ser Ile Glu Asp Trp Trp Arg Asn Glu Gln Phe Trp Val
915 920 925

Ile Gly Gly Thr Ser Ala His Leu Phe Ala Val Phe Gln Gly Leu Leu
930 935 940

```
Lys Val Leu Ala Gly Ile Asp Thr Asn Phe Thr Val Thr Ser Lys Ala
945                 950                 955                 960

Thr Asp Glu Asp Gly Asp Phe Ala Glu Leu Tyr Ile Phe Lys Trp Thr
                965                 970                 975

Ala Leu Leu Ile Pro Pro Thr Thr Val Leu Leu Val Asn Leu Ile Gly
            980                 985                 990

Ile Val Ala Gly Val Ser Tyr Ala Val Asn Ser Gly Tyr Gln Ser Trp
        995                 1000                1005

Gly Pro Leu Phe Gly Lys Leu Phe Phe Ala Leu Trp Val Ile Ala His
    1010                1015                1020

Leu Tyr Pro Phe Leu Lys Gly Leu Leu Gly Arg Gln Asn Arg Thr Pro
1025                1030                1035                1040

Thr Ile Val Ile Val Trp Ser Val Leu Leu Ala Ser Ile Phe Ser Leu
                1045                1050                1055

Leu Trp Val Arg Ile Asn Pro Phe Val Asp Ala Asn Pro Asn Ala Asn
            1060                1065                1070

Asn Phe Asn Gly Lys Gly Gly Val Phe
        1075                1080
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1741 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Oryza sativa

(vii) IMMEDIATE SOURCE:
(B) CLONE: S0542

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 101..1741

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTGCGGCCGC CGCGCATCTA GGCTTGCCGC GCGCGCGCGG ATCTGCGAGC TGCGTAGCCG 60

TTTCTCGCTG TGAGTGGAGG AGGAGGAGGA AGGGAGGAGG ATG GCG GCG AAC GCG 115
Met Ala Ala Asn Ala
1 5

GGG ATG GTG GCG GGA TCC CGC AAC CGG AAC GAG TTC GTC ATG ATC CGC 163
Gly Met Val Ala Gly Ser Arg Asn Arg Asn Glu Phe Val Met Ile Arg
10 15 20

CCC GAC GGC GAC GCG CCA CCG CCG GCT AAG CCA GGG AAG AGT GTG AAT 211
Pro Asp Gly Asp Ala Pro Pro Pro Ala Lys Pro Gly Lys Ser Val Asn
25 30 35

GGT CAG GTC TGC CAG ATT TGT GGC GAC ACT GTT GGC GTC TCG GCC ACC 259
Gly Gln Val Cys Gln Ile Cys Gly Asp Thr Val Gly Val Ser Ala Thr
40 45 50

GGC GAC GTC TTT GTT GCC TGC AAT GAG TGC GCC TTC CCG GTC TGC CGC 307
Gly Asp Val Phe Val Ala Cys Asn Glu Cys Ala Phe Pro Val Cys Arg
55 60 65

CCT TGC TAC GAG TAC GAA CGC AAG GAA GGG AAC CAG TGC TGC CCC CAG 355
Pro Cys Tyr Glu Tyr Glu Arg Lys Glu Gly Asn Gln Cys Cys Pro Gln
70 75 80 85

```
TGC AAG ACT AGA TAC AAG AGG CAC AAA GGT TGC CCT AGA GTT CAG GGC      403
Cys Lys Thr Arg Tyr Lys Arg His Lys Gly Cys Pro Arg Val Gln Gly
                 90                  95                 100

GAT GAG GAA GAA GAA GAT GTT GAT GAC CTG GAC AAT GAA TTC CAT TAT      451
Asp Glu Glu Glu Glu Asp Val Asp Asp Leu Asp Asn Glu Phe His Tyr
            105                 110                 115

AAG CAT GGC AAT GGC AAA GGT CCA GAG TGG CAG ATA CAG AGA CAG GGG      499
Lys His Gly Asn Gly Lys Gly Pro Glu Trp Gln Ile Gln Arg Gln Gly
        120                 125                 130

GAA GAT GTT GAC CTG TCT TCA TCT TCT CGC CAC GAA CAA CAT CGG ATT      547
Glu Asp Val Asp Leu Ser Ser Ser Ser Arg His Glu Gln His Arg Ile
    135                 140                 145

CCC CGT CTG ACA AGT GGG CAA CAG ATC TCA GGA GAG ATC CCT GAT GCT      595
Pro Arg Leu Thr Ser Gly Gln Gln Ile Ser Gly Glu Ile Pro Asp Ala
150                 155                 160                 165

TCC CCC GAT CGC CAT TCT ATC CGC AGC GGA ACA TCA AGC TAT GTT GAT      643
Ser Pro Asp Arg His Ser Ile Arg Ser Gly Thr Ser Ser Tyr Val Asp
                170                 175                 180

CCA AGT GTT CCA GTT CCT GTG AGG ATT GTG GAC CCC TCC AAG GAC TTG      691
Pro Ser Val Pro Val Pro Val Arg Ile Val Asp Pro Ser Lys Asp Leu
            185                 190                 195

AAT TCC TAT GGG ATT AAC AGT GTT GAC TGG CAA GAA AGA GTT GCC AGC      739
Asn Ser Tyr Gly Ile Asn Ser Val Asp Trp Gln Glu Arg Val Ala Ser
        200                 205                 210

TGG AGG AAC AAG CAG GAC AAA AAT ATG ATG CAG GTA GCT AAT AAA TAT      787
Trp Arg Asn Lys Gln Asp Lys Asn Met Met Gln Val Ala Asn Lys Tyr
    215                 220                 225

CCA GAG GCA AGA GGG GGA GAC ATG GAA GGG ACT GGT TCA AAT GGT GAA      835
Pro Glu Ala Arg Gly Gly Asp Met Glu Gly Thr Gly Ser Asn Gly Glu
230                 235                 240                 245
```

```
GAT ATC CAA ATG GTT GAT GAT GCA CGT CTA CCT CTG AGC CGC ATA GTG      883
Asp Ile Gln Met Val Asp Asp Ala Arg Leu Pro Leu Ser Arg Ile Val
                250                 255                 260

CCT ATC CCT TCA AAC CAG CTC AAC CTT TAC CGG ATT GTT ATC ATT CTC      931
Pro Ile Pro Ser Asn Gln Leu Asn Leu Tyr Arg Ile Val Ile Ile Leu
            265                 270                 275

CGT CTT ATC ATC CTG ATG TTC TTC TTC CAA TAT CGT GTC ACT CAT CCA      979
Arg Leu Ile Ile Leu Met Phe Phe Phe Gln Tyr Arg Val Thr His Pro
        280                 285                 290

GTG CGG GAT GCT TAT GGA TTG TGG CTA GTA TCT GTT ATC TGT GAA ATT     1027
Val Arg Asp Ala Tyr Gly Leu Trp Leu Val Ser Val Ile Cys Glu Ile
    295                 300                 305

TGG TTG CCC TTA TCC TGG CTC CTA GAT CAA TTC CCA AAG TGG TAC CCG     1075
Trp Leu Pro Leu Ser Trp Leu Leu Asp Gln Phe Pro Lys Trp Tyr Pro
310                 315                 320                 325

ATA AAC CGT GAA ACA TAC CTT GAC AGG CTT GCA TTG AGA TAT GAT AGG     1123
Ile Asn Arg Glu Thr Tyr Leu Asp Arg Leu Ala Leu Arg Tyr Asp Arg
                330                 335                 340

GAG GGA GAG CCA TCA CAG CTT GCT CCC ATT GAT GTC TTT GTC AGT ACG     1171
Glu Gly Glu Pro Ser Gln Leu Ala Pro Ile Asp Val Phe Val Ser Thr
            345                 350                 355

GTG GAT CCA CTA AAG GAA CCT CCT CTG ATC ACA GCA AAC ACT GTT TTG     1219
Val Asp Pro Leu Lys Glu Pro Pro Leu Ile Thr Ala Asn Thr Val Leu
        360                 365                 370

TCC ATT CTG GCT GTG GAT TAC CCT GTT GAC AAA GTG TCA TGC TAT GTT     1267
Ser Ile Leu Ala Val Asp Tyr Pro Val Asp Lys Val Ser Cys Tyr Val
    375                 380                 385

TCT GAC GAT GGT TCA GCT ATG TTA ACT TTT GAG GCT CTG TCA GAA ACT     1315
Ser Asp Asp Gly Ser Ala Met Leu Thr Phe Glu Ala Leu Ser Glu Thr
390                 395                 400                 405
```

```
GCA GAA TTT GCT AGG AAG TGG GTT CCG TTT TGC AAG AAG CAC AAT ATT      1363
Ala Glu Phe Ala Arg Lys Trp Val Pro Phe Cys Lys Lys His Asn Ile
                410                 415                 420

GAA CCA CGA GCT CCA GAG TTT TAC TTT GCT CAA AAA ATA GAT TAC CTG      1411
Glu Pro Arg Ala Pro Glu Phe Tyr Phe Ala Gln Lys Ile Asp Tyr Leu
            425                 430                 435

AAG GAC AAA ATC CAA CCT TCC TTT GTT AAA GAA AGG CGG GCA ATG AAG      1459
Lys Asp Lys Ile Gln Pro Ser Phe Val Lys Glu Arg Arg Ala Met Lys
        440                 445                 450

AGA GAG TAT GAA GAA TTC AAG GTA CGG ATC AAT GCT CTT GTT GCG AAG      1507
Arg Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn Ala Leu Val Ala Lys
    455                 460                 465

GCA CAA AAA GTA CCT GAA GAG GGG TGG ACC ATG GCT GAT GGC ACT GCT      1555
Ala Gln Lys Val Pro Glu Glu Gly Trp Thr Met Ala Asp Gly Thr Ala
470                 475                 480                 485

TGG CCT GGG AAT AAC CCA AGG GAT CAC CCT GGC ATG ATT CAG GTG TTC      1603
Trp Pro Gly Asn Asn Pro Arg Asp His Pro Gly Met Ile Gln Val Phe
                490                 495                 500

TTG GGG CAC AGT GGT GGG CTT GAC ACT GAT GGT AAC GAG TTG CCA CGG      1651
Leu Gly His Ser Gly Gly Leu Asp Thr Asp Gly Asn Glu Leu Pro Arg
            505                 510                 515

CTT GTC TAC GTC TCT CGT GAA AAG AGG CCA GGA TTC CAG CAT CAC AAG      1699
Leu Val Tyr Val Ser Arg Glu Lys Arg Pro Gly Phe Gln His His Lys
        520                 525                 530

AAG GCT GGT GCA ATG AAT GCA TTG ATT CGT GTA TCT GCT GTG              1741
Lys Ala Gly Ala Met Asn Ala Leu Ile Arg Val Ser Ala Val
    535                 540                 545
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 547 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Ala Ala Asn Ala Gly Met Val Ala Gly Ser Arg Asn Arg Asn Glu
 1               5                  10                  15

Phe Val Met Ile Arg Pro Asp Gly Asp Ala Pro Pro Pro Ala Lys Pro
            20                  25                  30

Gly Lys Ser Val Asn Gly Gln Val Cys Gln Ile Cys Gly Asp Thr Val
        35                  40                  45

Gly Val Ser Ala Thr Gly Asp Val Phe Val Ala Cys Asn Glu Cys Ala
    50                  55                  60

Phe Pro Val Cys Arg Pro Cys Tyr Glu Tyr Glu Arg Lys Glu Gly Asn
65                  70                  75                  80

Gln Cys Cys Pro Gln Cys Lys Thr Arg Tyr Lys Arg His Lys Gly Cys
                85                  90                  95

Pro Arg Val Gln Gly Asp Glu Glu Glu Glu Asp Val Asp Asp Leu Asp
            100                 105                 110

Asn Glu Phe His Tyr Lys His Gly Asn Gly Lys Gly Pro Glu Trp Gln
        115                 120                 125

Ile Gln Arg Gln Gly Glu Asp Val Asp Leu Ser Ser Ser Ser Arg His
    130                 135                 140

Glu Gln His Arg Ile Pro Arg Leu Thr Ser Gly Gln Gln Ile Ser Gly
145                 150                 155                 160
```

```
Glu Ile Pro Asp Ala Ser Pro Asp Arg His Ser Ile Arg Ser Gly Thr
                165                 170                 175

Ser Ser Tyr Val Asp Pro Ser Val Pro Val Pro Val Arg Ile Val Asp
            180                 185                 190

Pro Ser Lys Asp Leu Asn Ser Tyr Gly Ile Asn Ser Val Asp Trp Gln
        195                 200                 205

Glu Arg Val Ala Ser Trp Arg Asn Lys Gln Asp Lys Asn Met Met Gln
    210                 215                 220

Val Ala Asn Lys Tyr Pro Glu Ala Arg Gly Gly Asp Met Glu Gly Thr
225                 230                 235                 240

Gly Ser Asn Gly Glu Asp Ile Gln Met Val Asp Asp Ala Arg Leu Pro
                245                 250                 255

Leu Ser Arg Ile Val Pro Ile Pro Ser Asn Gln Leu Asn Leu Tyr Arg
            260                 265                 270

Ile Val Ile Ile Leu Arg Leu Ile Ile Leu Met Phe Phe Phe Gln Tyr
        275                 280                 285

Arg Val Thr His Pro Val Arg Asp Ala Tyr Gly Leu Trp Leu Val Ser
    290                 295                 300

Val Ile Cys Glu Ile Trp Leu Pro Leu Ser Trp Leu Leu Asp Gln Phe
305                 310                 315                 320

Pro Lys Trp Tyr Pro Ile Asn Arg Glu Thr Tyr Leu Asp Arg Leu Ala
                325                 330                 335

Leu Arg Tyr Asp Arg Glu Gly Glu Pro Ser Gln Leu Ala Pro Ile Asp
            340                 345                 350

Val Phe Val Ser Thr Val Asp Pro Leu Lys Glu Pro Pro Leu Ile Thr
        355                 360                 365
```

```
Ala Asn Thr Val Leu Ser Ile Leu Ala Val Asp Tyr Pro Val Asp Lys
    370                 375                 380

Val Ser Cys Tyr Val Ser Asp Asp Gly Ser Ala Met Leu Thr Phe Glu
385                 390                 395                 400

Ala Leu Ser Glu Thr Ala Glu Phe Ala Arg Lys Trp Val Pro Phe Cys
                405                 410                 415

Lys Lys His Asn Ile Glu Pro Arg Ala Pro Glu Phe Tyr Phe Ala Gln
            420                 425                 430

Lys Ile Asp Tyr Leu Lys Asp Lys Ile Gln Pro Ser Phe Val Lys Glu
        435                 440                 445

Arg Arg Ala Met Lys Arg Glu Tyr Glu Glu Phe Lys Val Arg Ile Asn
    450                 455                 460

Ala Leu Val Ala Lys Ala Gln Lys Val Pro Glu Glu Gly Trp Thr Met
465                 470                 475                 480

Ala Asp Gly Thr Ala Trp Pro Gly Asn Asn Pro Arg Asp His Pro Gly
                485                 490                 495

Met Ile Gln Val Phe Leu Gly His Ser Gly Gly Leu Asp Thr Asp Gly
            500                 505                 510

Asn Glu Leu Pro Arg Leu Val Tyr Val Ser Arg Glu Lys Arg Pro Gly
        515                 520                 525

Phe Gln His His Lys Lys Ala Gly Ala Met Asn Ala Leu Ile Arg Val
    530                 535                 540

Ser Ala Val
545
```

CLAIMS:

1. An isolated nucleic acid molecule which encodes a polypeptide of the cellulose biosynthetic pathway or a homologue, analogue or derivative thereof or a complementary sequence thereto, wherein said polypeptide is capable of producing cellulose and/or β-1,4-glucan and/or an intermediate between cellulose and a β-1,4-glucan polymer.

2. The isolated nucleic acid molecule according to claim 1 wherein the polypeptide is cellulose synthase or a catalytic subunit thereof.

3. The isolated nucleic acid molecule according to claim 1 or 2, derived from a prokaryote.

4. The isolated nucleic acid molecule according to claim 3, wherein the prokaryote is a bacterium other than *Agrobacterium tumefaciens*, *Acetobacter pasteurianus* or *Acetobacter xylinum*.

5. The isolated nucleic acid molecule according to claim 1 or 2, derived from a eukaryote.

6. The isolated nucleic acid molecule according to claim 5, wherein the eukaryote is a plant or fungus.

7. The isolated nucleic acid molecule according to claim 6, wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), wheat, barley, maize, *Brassica ssp.*, *Eucalyptus ssp.*, hemp, jute, flax, *Pinus ssp.*, *Populus ssp.*, and *Picea spp.*, amongst others.

8. The isolated nucleic acid molecule according to claim 2 wherein the cellulose synthase or catalytic subunit thereof is the *Arabidopsis thaliana* RSW1 polypeptide.

9. The isolated nucleic acid molecule according to any one of claims 1 to 8, comprising a sequence of nucleotides which is at least 40% identical to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereof.

10. The isolated nucleic acid molecule according to claim 9, wherein the percentage identity to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereof is at least 60%.

11. The isolated nucleic acid molecule according to claim 9, wherein the percentage identity to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereof is at least 80%.

12. An isolated nucleic acid molecule which comprises a sequence of nucleotides substantially as set forth in any one of SEQ ID NOs:3, 4, 5, 7, 9 or 11 or a homologue, analogue or derivative thereof or a complementary sequence thereto.

13. The isolated nucleic acid molecule according to any one of claims 1 to 12 , wherein said nucleic acid molecule hybridizes under at least low stringency conditions to at least 20 contiguous nucleotides of any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereto.

14. An isolated nucleic acid molecule which encodes a polypeptide which is capable of cellulose and/or β-1,4- glucan biosynthesis in a plant cell, fungal cell, insect cell. animal cell, yeast cell or bacterial cell when expressed therein.

15. The isolated nucleic acid molecule according to claim 14, wherein the polypeptide is cellulose synthase or a catalytic subunit thereof.

16. The isolated nucleic acid molecule according to claim 14 or 15, derived from a

prokaryote.

17. The isolated nucleic acid molecule according to claim 16, wherein the prokaryote is a bacterium other than *Agrobacterium tumefaciens*, *Acetobacter pasteurianus* or *Acetobacter xylinum*.

18. The isolated nucleic acid molecule according to claim 14 or 15, derived from a eukaryote.

19. The isolated nucleic acid molecule according to claim 18, wherein the eukaryote is a plant or fungus.

20. The isolated nucleic acid molecule according to claim 19, wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), wheat, barley, maize, *Brassica ssp.*, *Eucalyptus ssp.*, hemp, jute, flax, *Pinus ssp.*, *Populus ssp.*, and *Picea spp.*, amongst others.

21. The isolated nucleic acid molecule according to claim 20, wherein the cellulose synthase or catalytic subunit thereof is the *Arabidopsis thaliana* RSW1 polypeptide.

22. The isolated nucleic acid molecule according to any one of claims 14 to 21, comprising a sequence of nucleotides which is at least 40% identical to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereto.

23. The isolated nucleic acid molecule according to claim 22, wherein the percentage identity to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence thereof is at least 60%.

24. The isolated nucleic acid molecule according to claim 22, wherein the percentage identity to any one of SEQ ID NOs:1, 3, 4, 5, 7, 9, 11 or 13 or a complementary sequence

thereof is at least 80%.

25. The isolated nucleic acid molecule according to claim 22 , comprising the sequence of nucleotides substantially as set forth in any one of SEQ ID NOs:3, 4, 5, 7, 9 or 11 or a homologue, analogue or derivative thereof or a complementary sequence thereto.

26. An isolated nucleic acid molecule which encodes or is complementary to a nucleic acid molecule which encodes a polypeptide capable of cellulose and/or β-1,4-glucan biosynthesis wherein said polypeptide comprises a sequence of amino acids which is at least 40% identical to any one of SEQ ID Nos:2, 6, 8, 10, 12 or 14.

27. The isolated nucleic acid molecule according to claim 26, wherein the percentage identity to any one of SEQ ID Nos:2, 6, 8, 10, 12 or 14 is at least 60%.

28. The isolated nucleic acid molecule according to claim 27, wherein the percentage identity to any one of SEQ ID Nos:2, 6, 8, 10, 12 or 14 is at least 80%.

29. The isolated nucleic acid molecule according to claim 26, wherein the polypeptide comprises a sequence of amino acids substantially as set forth in any one of SEQ ID Nos:2, 6, 8, 10, 12 or 14.

30. A genetic construct which comprises the isolated nucleic acid molecule according to any one of claims 1 to 29.

31. A genetic construct which comprises the isolated nucleic acid molecule according to any one of claims 1 to 29 operably connected to a promoter sequence.

32. The genetic construct according to claim 31, wherein the nucleic acid molecule is operably connected to the promoter sequence in the sense orientation such that RNA which encodes a polypeptide capable of cellulose and/or β-1,4-glucan biosynthesis or a homologue,

analogue or derivative thereof is produced when said nucleic acid molecule is expressed.

33. The genetic construct according to claim 31, wherein the nucleic acid molecule is operably connected to the promoter sequence in the antisense orientation such that RNA which is complementary to RNA which encodes a polypeptide capable of cellulose and/or β-1,4-glucan biosynthesis or a homologue, analogue or derivative thereof, is produced when said nucleic acid molecule is expressed.

34. The genetic construct according to claim 33, wherein the nucleic acid molecule encodes an antisense or ribozyme molecule.

35. The genetic construct according to any one of claims 31 to 34, wherein the promoter is the CaMV 35S promoter.

36. The genetic construct according to any one of claims 31 to 34, wherein the promoter is the *Arabidopsis thaliana RSW1* gene promoter.

37. A method of increasing the level of cellulose in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 therein, in the sense orientation, for a time and under conditions at least sufficient to produce or increase expression of the polypeptide encoded therefor.

38. The method according to claim 37, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

39. The method according to claim 38, wherein the cell is a prokaryotic cell.

40. The method according to claim 38, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

41. The method according to claim 40, wherein the cell, tissue, organ or organism is a plant, fungal, insect, animal or yeast cell, tissue, organ or organism.

42. The method according to claim 41, wherein the cell, tissue, organ or organism is a plant cell, tissue, organ or organism.

43. The method according to claim 42 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

44. A method of reducing the level of non-crystalline β-1,4-glucan in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 therein, in the sense orientation, for a time and under conditions at least sufficient to produce or increase expression of the polypeptide encoded therefor.

45. The method according to claim 44, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

46. The method according to claim 44, wherein the cell is a prokaryotic cell.

47. The method according to claim 44, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

48. The method according to claim 47, wherein the cell, tissue, organ or organism is a plant, fungal, insect, animal or yeast cell, tissue, organ or organism.

49. The method according to claim 48, wherein the cell, tissue, organ or organism is a plant cell, tissue, organ or organism.

50. The method according to claim 50 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

51. A method of reducing the level of starch in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 therein, in the sense orientation, for a time and under conditions at least sufficient to produce or increase expression of the polypeptide encoded therefor.

52. The method according to claim 50, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

53. The method according to claim 51, wherein the cell is a prokaryotic cell.

54. The method according to claim 53, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

55. The method according to claim 54, wherein the eukaryote is a plant, fungus, insect, animal or yeast.

56. The method according to claim 55, wherein the eukaryote is a plant.

57. The method according to claim 56 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

58. A method of reducing the level of cellulose in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of

claims 1 to 29 therein, in the antisense orientation, for a time and under conditions at least sufficient to prevent or reduce the expression of the polypeptide encoded therefor.

59. The method according to claim 58, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

60. The method according to claims 58 or 59, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

61. The method according to claim 60, wherein the eukaryote is a plant, fungus, insect, animal or yeast.

62. The method according to claim 61, wherein the eukaryote is a plant.

63. The method according to claim 62 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

64. A method of increasing the level of non-crystalline β-1,4-glucan in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 therein, in the antisense orientation, for a time and under conditions at least sufficient to prevent or reduce the expression of the polypeptide encoded therefor.

65. The method according to claim 64, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

66. The method according to claims 64 or 65, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

67. The method according to claim 66, wherein the eukaryote is a plant, fungus, insect, animal or yeast.

68. The method according to claim 67, wherein the eukaryote is a plant.

69. The method according to claim 68 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

70. A method of increasing the level of starch in a cell, tissue, organ or organism, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 therein, in the antisense orientation, for a time and under conditions at least sufficient to prevent or reduce the expression of the polypeptide encoded therefor.

71. The method according to claim 70, comprising the additional first step of transforming the cell, tissue, organ or organism with the isolated nucleic acid molecule.

72. The method according to claims 70 or 71, wherein the cell, tissue, organ or organism is a eukaryotic cell, tissue, organ or organism.

73. The method according to claim 72, wherein the eukaryote is a plant, fungus, insect, animal or yeast.

74. The method according to claim 73, wherein the eukaryote is a plant.

75. The method according to claim 74 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

76. A method of producing a recombinant enzymatically active polypeptide which is capable of synthesizing cellulose and/or β-1,4-glucan and/or an intermediate between cellulose and β-1,4-glucan in a cell, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 or a homologue, analogue or derivative thereof in said cell for a time and under conditions sufficient for the polypeptide encoded therefor to be produced.

77. The method according to claim 76, comprising the additional first step of transforming the cell with the isolated nucleic acid molecule according to any one of claims 1 to 29 or the genetic construct according to any one of claims 11 to 15.

78. A recombinant polypeptide produced according to the method defined by claim 76 or 77.

79. The recombinant cellulose biosynthetic polypeptide according to claim 78, further defined as a recombinant cellulose synthase or catalytically active subunit thereof.

80. A recombinant cellulose biosynthetic polypeptide capable of cellulose and/or β-1,4-glucan production and comprising a sequence of amino acids set forth in any one of SEQ ID Nos: 2, 6, 8, 10, 12 or 14 or a homologue, analogue or derivative thereof which is at least 40% identical thereto.

81. The recombinant cellulose biosynthetic polypeptide according to claim 80, wherein the percentage identity to any one of SEQ ID Nos: 2, 6, 8, 10, 12 or 14 is at least 60%.

82. The recombinant cellulose biosynthetic polypeptide according to claim 81, wherein the percentage identity to any one of SEQ ID Nos: 2, 6, 8, 10, 12 or 14 is at least 80%.

83. The recombinant cellulose biosynthetic polypeptide according to claim 82, comprising a sequence of amino acids substantially as set forth in any one of SEQ ID Nos: 2, 6, 8, 10,

12 or 14.

84. A method of altering the mechanical properties of a cell wall, said method comprising expressing the isolated nucleic acid molecule according to any one of claims 1 to 29 in the antisense orientation in said cell for a time and under conditions sufficient for the level of non-crystalline β-1,4-glucan to increase in said cell.

85. The method according to claim 84 wherein the non-crystalline β-1,4-glucan is cross-linked to cellulose microfibrils.

86. The method according to claim 84 or 85 wherein the cell wall normally has a high ratio of cellulose to hemicelluloses.

87. The method according to any one of claims 84 to 86, wherein the nucleic acid molecule expressed in the antisense orientation is contained within an antisense molecule or ribozyme molecule.

88. The method according to any one of claims 84 to 87, wherein the cell wall is a plant cell wall.

89. The method according to claim 88, wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

90. An antibody molecule which binds to the recombinant polypeptide according to any one of claims 78 to 83 or a homologue, analogue or derivative thereof.

91. A transgenic plant transformed with the isolated nucleic acid molecule according to any one of claims 1 to 29 or a genetic construct according to any one of claims 30 to 36.

92. The transgenic plant according to claim 91, wherein said plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

93. Use of an isolated nucleic acid molecule according to any one of claims 1 to 29 to modify the cellulose content of a cell.

94. Use according to claim 93, wherein if the nucleic acid molecule according to any one of claims 1 to 29 is expressed in the sense orientation in said cell, the level of cellulose therein is increased.

95. Use according to claim 93, wherein if the nucleic acid molecule according to any one of claims 1 to 29 is expressed in the antisense orientation in said cell, the level of cellulose therein is decreased.

96. Use according to claim 95, wherein said cell is further characterised by increased non-crystalline β-1,4-glucan content and/or starch content.

97. Use according to claim 95 or 96, wherein said cell is further characterised by increased cross-linking of non-crystalline β-1,4-glucan to cellulose.

98. Use according to any one of claims 93 to 97, wherein the cell is a plant cell.

99. Use according to claim 98 wherein the plant is selected from the list comprising *Arabidopsis thaliana*, *Gossypium hirsutum* (cotton), *Oryza sativa* (rice), *Eucalyptus ssp.*, *Brassica ssp.*, wheat, barley, maize, hemp, jute, flax, and woody plants such as *Pinus ssp.*, *Populus ssp.*, *Picea spp.*, amongst others.

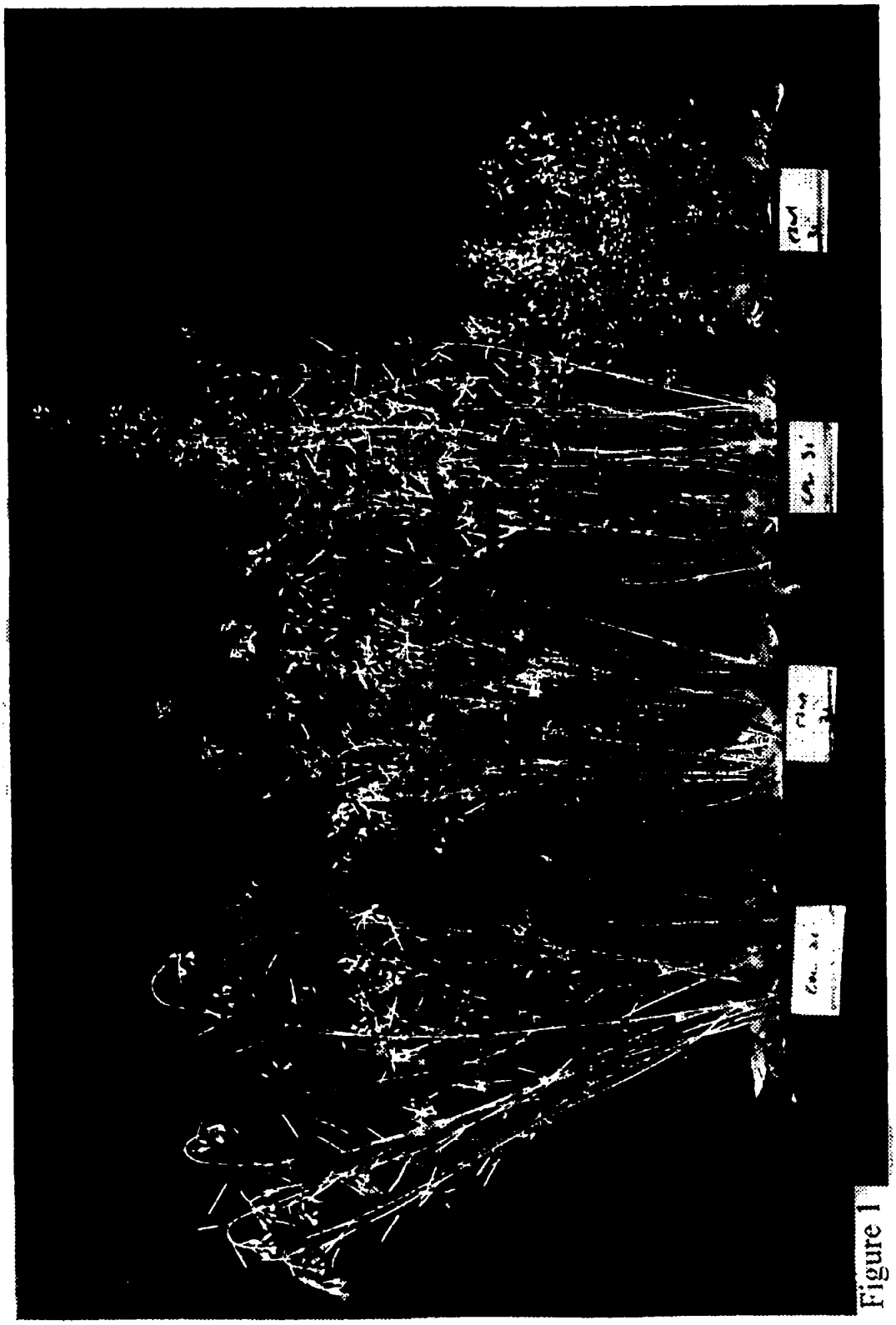


Figure 1

Figure 2

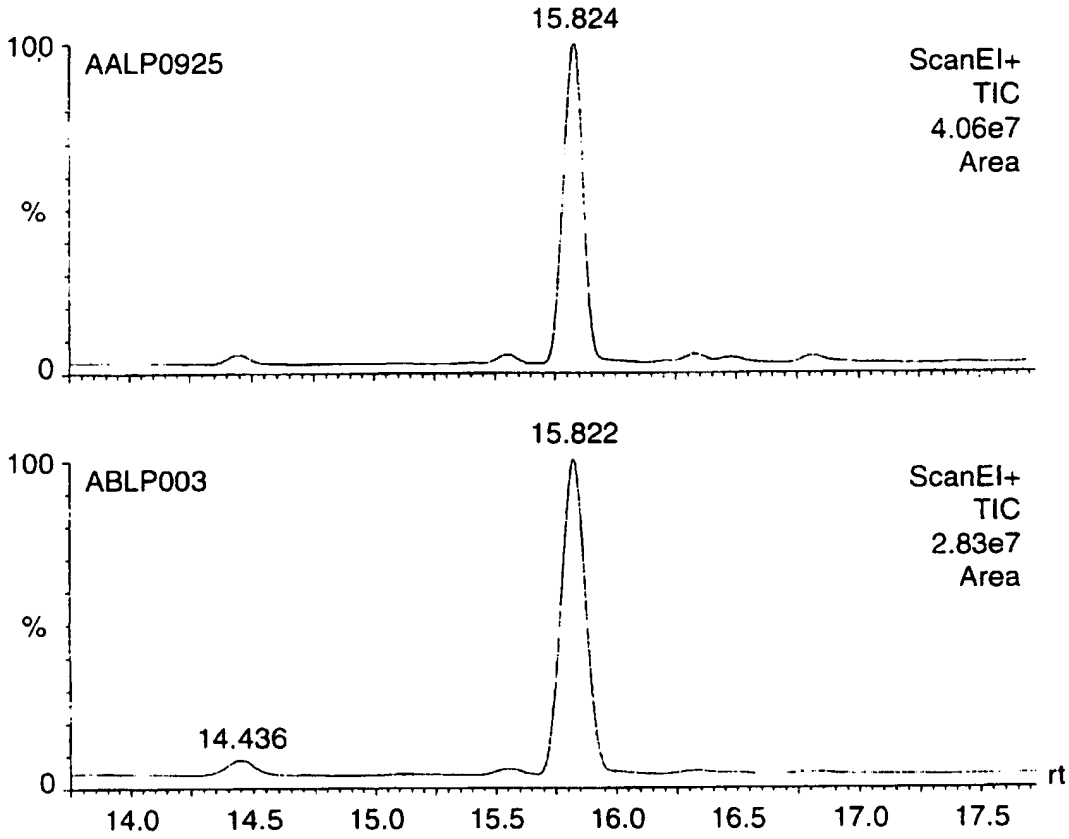

15.824
100
AALP0925
ScanEI+
TIC
4.06e7
Area
%
0
15.822
100
ABLP003
ScanEI+
TIC
2.83e7
Area
%
14.436
0
rt
14.0
14.5
15.0
15.5
16.0
16.5
17.0
17.5

Figure 3

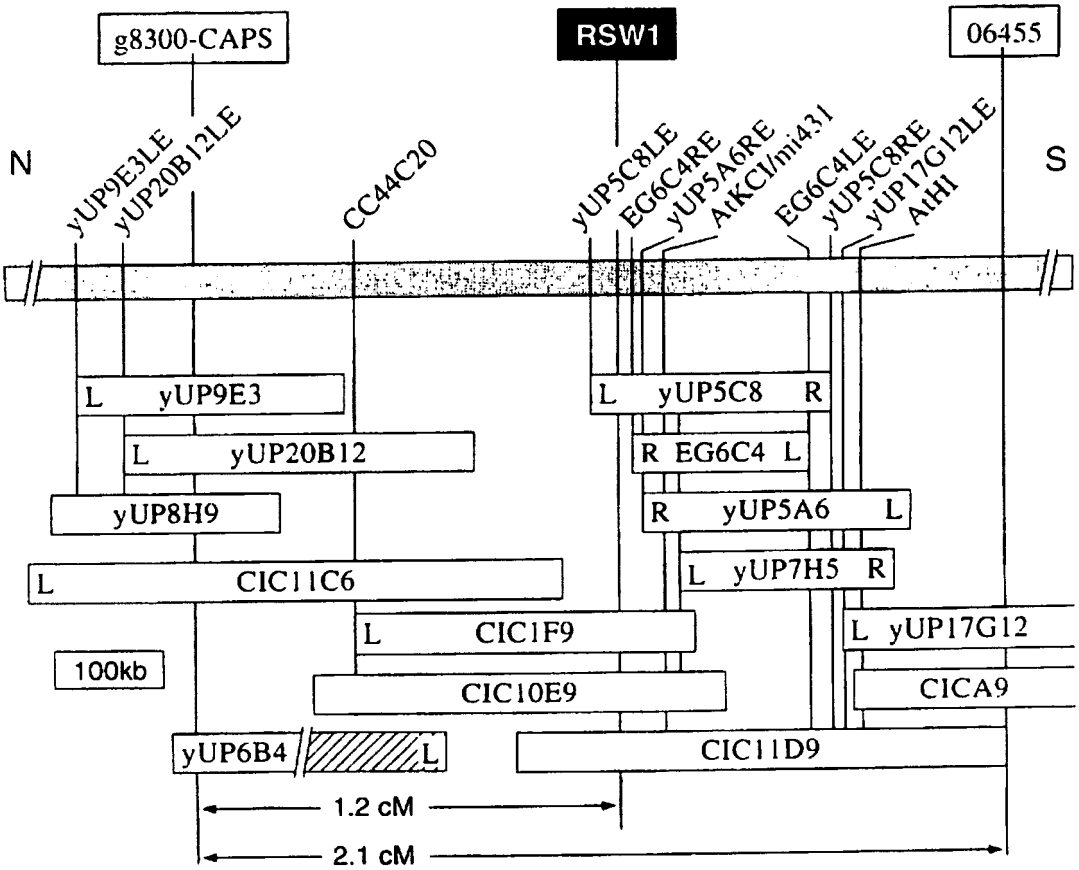

g8300-CAPS
RSW1
06455
N
S
yUP9E3LE
yUP20B12LE
CC44C20
yUP5C8LE
EG6C4RE
yUP5A6RE
AtKCI/mi431
EG6C4LE
yUP5C8RE
yUP17G12LE
AtHI
L yUP9E3
L yUP20B12
yUP8H9
L CIC11C6
L yUP5C8 R
R EG6C4 L
R yUP5A6 L
L yUP7H5 R
L CIC1F9
L yUP17G12
100kb
CIC10E9
CICA9
yUP6B4
L
CIC11D9
1.2 cM
2.1 cM

Figure 4

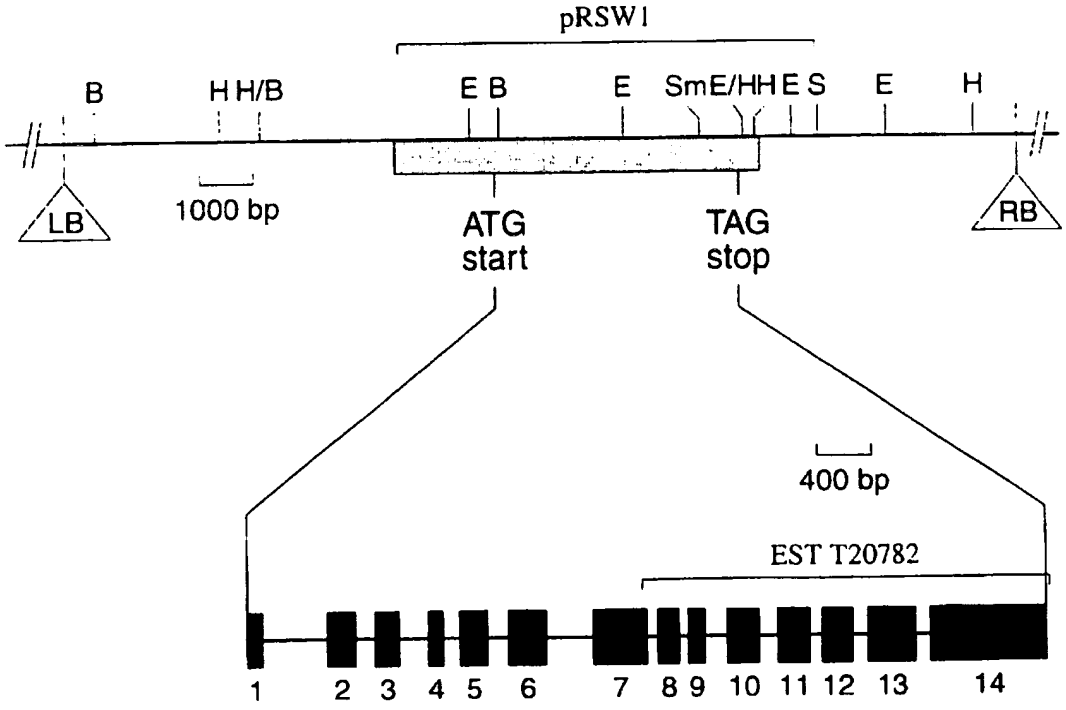

pRSW1
B
H H/B
E B
E
SmE/HH E S
E
H
LB
1000 bp
ATG
start
TAG
stop
RB
400 bp
EST T20782
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 5

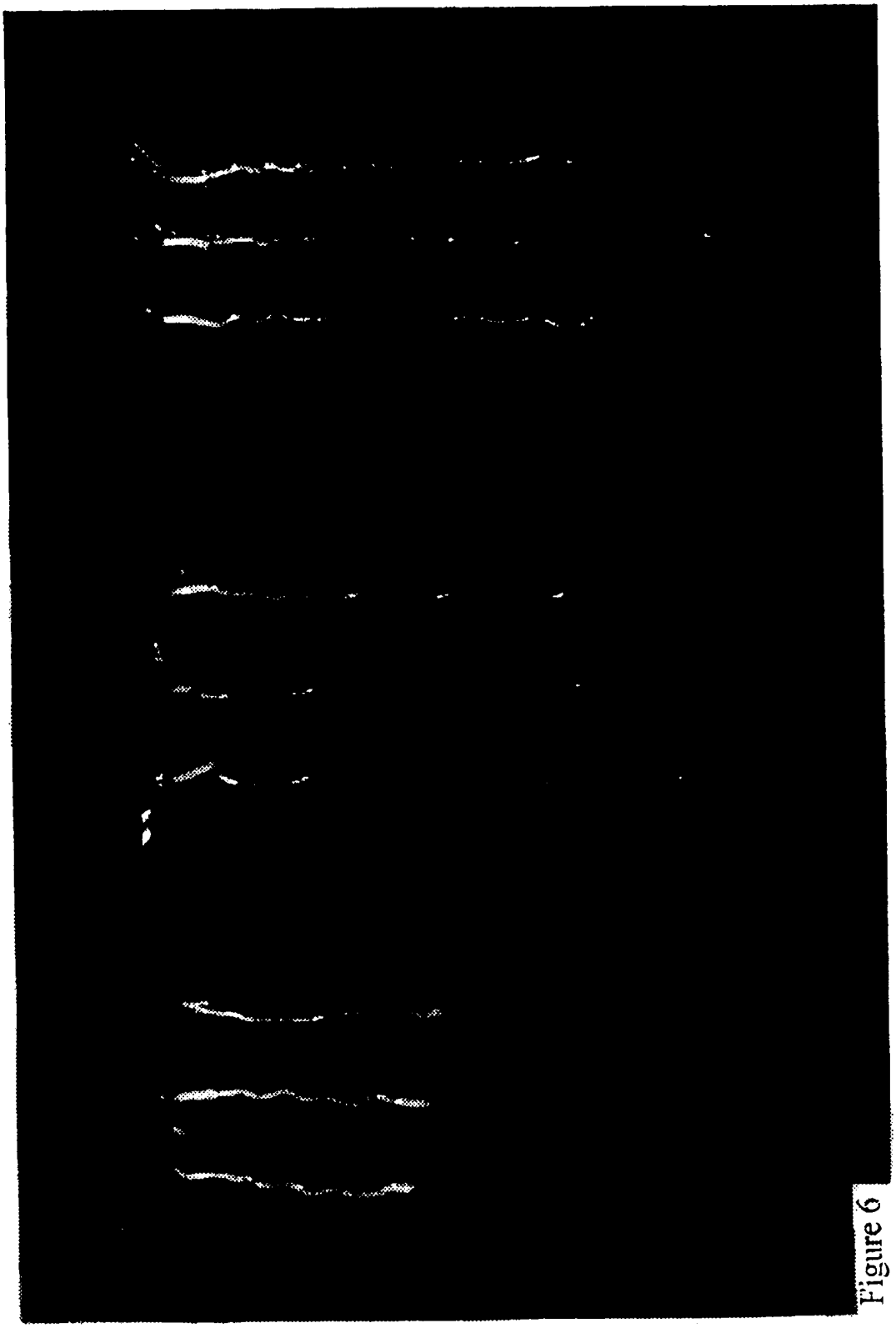


Figure 6

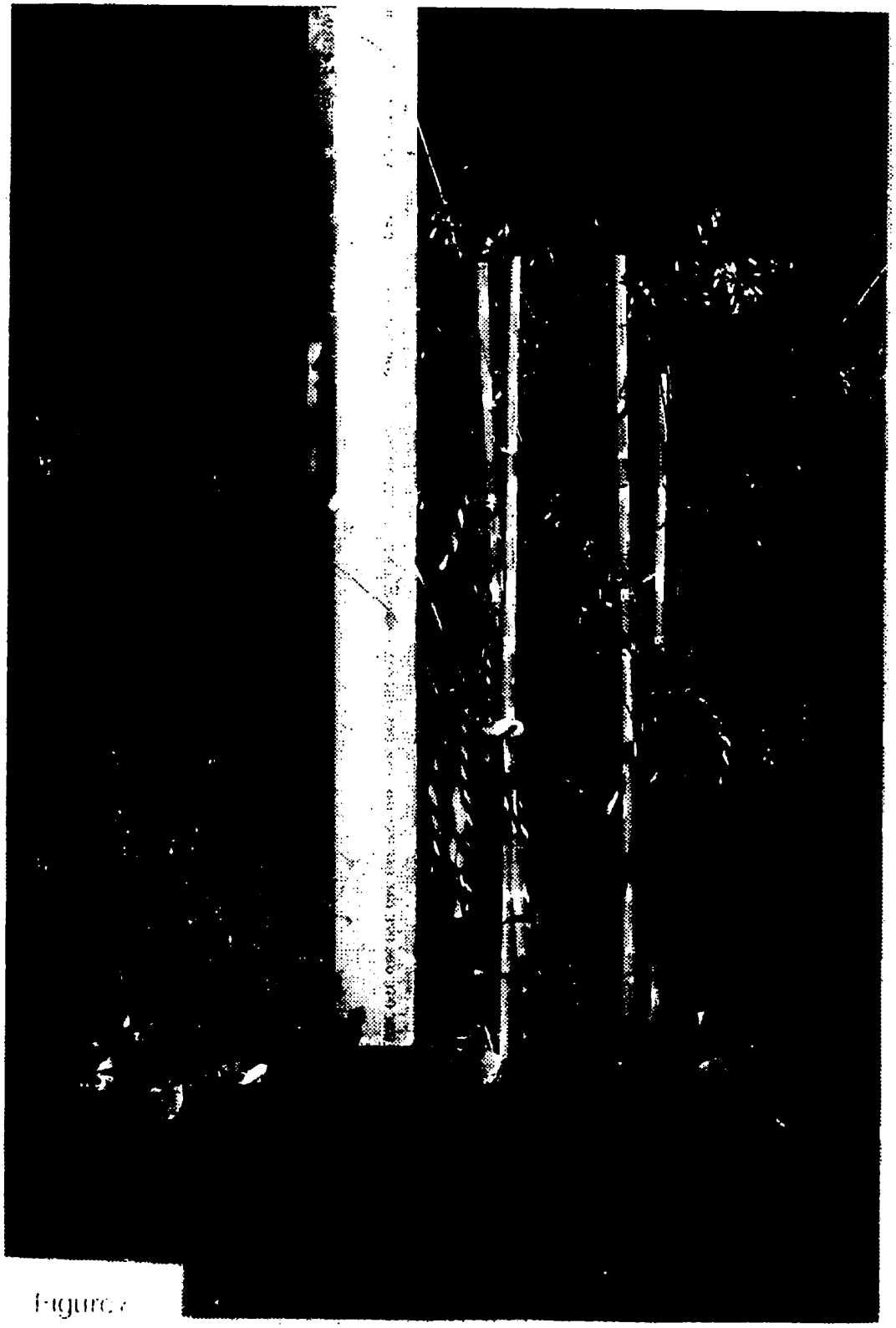


Figure 7

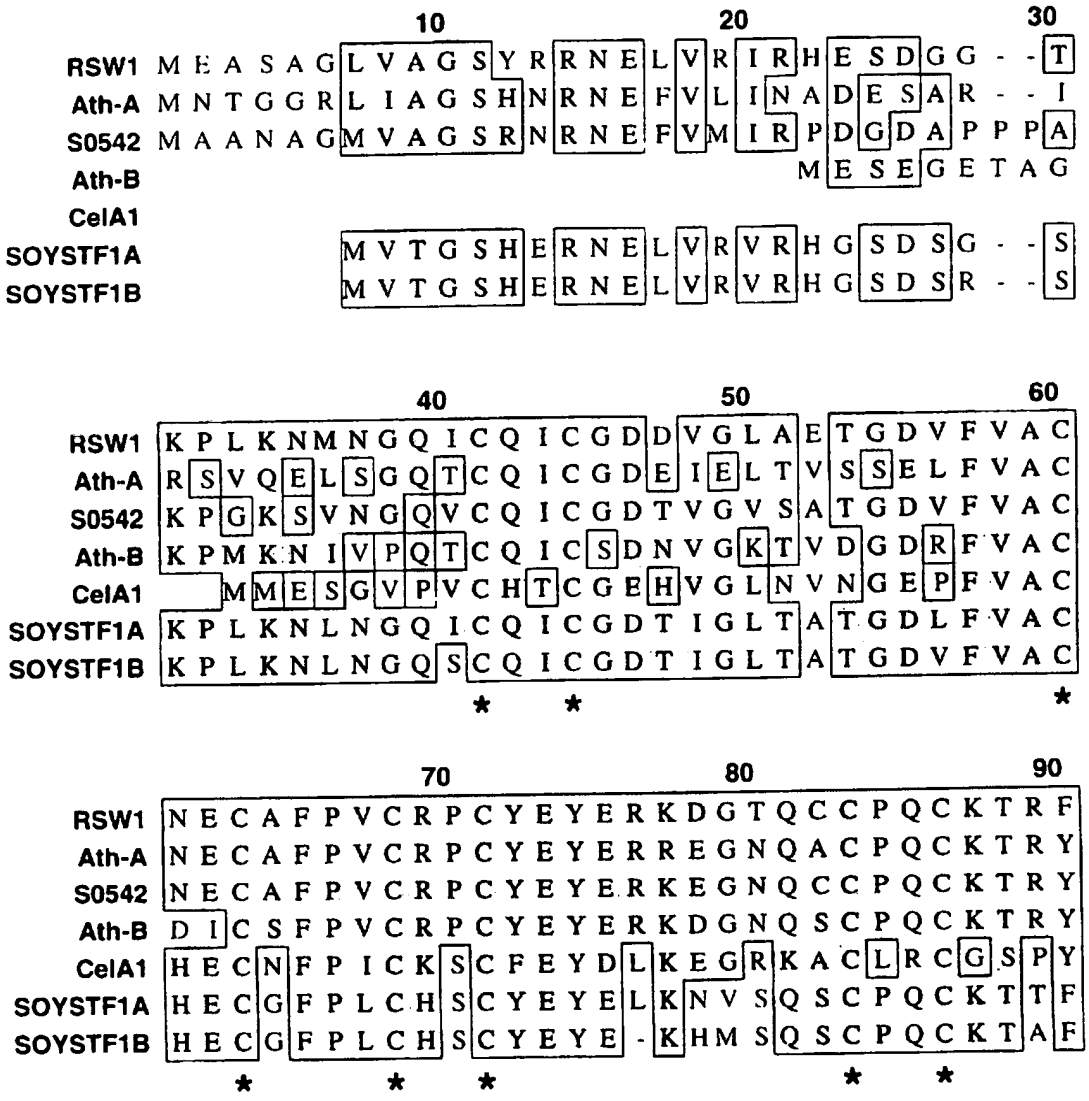
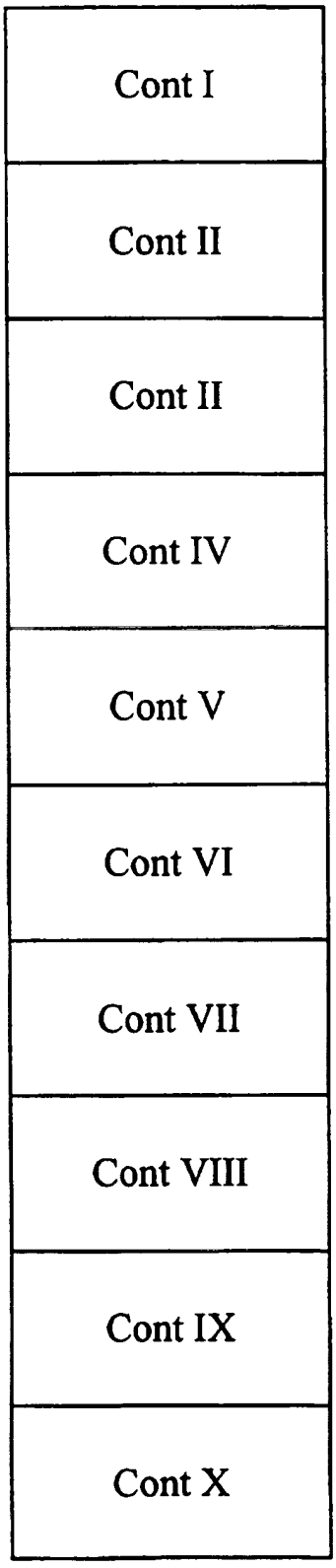

10
20
30
RSW1 M E A S A G L V A G S Y R R N E L V R I R H E S D G G - - T
Ath-A M N T G G R L I A G S H N R N E F V L I N A D E S A R - - I
S0542 M A A N A G M V A G S R N R N E F V M I R P D G D A P P P A
Ath-B M E S E G E T A G
CelA1
SOYSTF1A M V T G S H E R N E L V R V R H G S D S G - - S
SOYSTF1B M V T G S H E R N E L V R V R H G S D S R - - S
40
50
60
RSW1 K P L K N M N G Q I C Q I C G D D V G L A E T G D V F V A C
Ath-A R S V Q E L S G Q T C Q I C G D E I E L T V S S E L F V A C
S0542 K P G K S V N G Q V C Q I C G D T V G V S A T G D V F V A C
Ath-B K P M K N I V P Q T C Q I C S D N V G K T V D G D R F V A C
CelA1 M M E S G V P V C H T C G E H V G L N V N G E P F V A C
SOYSTF1A K P L K N L N G Q I C Q I C G D T I G L T A T G D L F V A C
SOYSTF1B K P L K N L N G Q S C Q I C G D T I G L T A T G D V F V A C
70
80
90
RSW1 N E C A F P V C R P C Y E Y E R K D G T Q C C P Q C K T R F
Ath-A N E C A F P V C R P C Y E Y E R R E G N Q A C P Q C K T R Y
S0542 N E C A F P V C R P C Y E Y E R K E G N Q C C P Q C K T R Y
Ath-B D I C S F P V C R P C Y E Y E R K D G N Q S C P Q C K T R Y
CelA1 H E C N F P I C K S C F E Y D L K E G R K A C L R C G S P Y
SOYSTF1A H E C G F P L C H S C Y E Y E L K N V S Q S C P Q C K T T F
SOYSTF1B H E C G F P L C H S C Y E Y E - K H M S Q S C P Q C K T A F

Figure 8


Cont I
Cont II
Cont II
Cont IV
Cont V
Cont VI
Cont VII
Cont VIII
Cont IX
Cont X

```
                10        20        30        40        50        60
RSW1    MEASAGLVAGSYRRNELVRIRHESDGG..TKPLKNMNGQICQICGDDVGLAETGDVFVAC
Ath-A   MNTGGRLIAGSHNRNEFVLINADESAR..IRSVQELSGQTCQICGDEIELTVSSELFVAC
S0542   MAANAGMVAGSRNRNEFVMIRPDGDAPPPAKPGKSVNGQVCQICGDTVGVSATGDVFVAC
Ath-B                           MESEGETAGKPMKNIVPQTCQICSDNVGKTVDGDRFVAC
CelA1                                  MMESGVPVCHTCGEHVGLNVNGEPFVAC
CelA2
D48636

                70        80        90        100       110       120
RSW1    NECAFPVCRPCYEYERKDGTQCCPQCKTRFRRHRGSPRVEGDEDEDDVDDIENEFNYAQG
Ath-A   NECAFPVCRPCYEYERREGNQACPQCKTRYKRIKGSPRVDGDDEEEEDIDDLEYEFDHGM
S0542   NECAFPVCRPCYEYERKEGNQCCPQCKTRYKRHKGCPRVQGDEEEEDVDDLDNEFHYKHG
Ath-B   DICSFPVCRPCYEYERKDGNQSCPQCKTRYKRLKGSPAIPGDKDEDGLADEGTVEFNYPQ
CelA1   HECNFPICKSCFEYDLKEGRKACLRCGSPYDENLLDDVEKATGDQSTMAAHLNKSQDVGI
CelA2
D48636
```

FIGURE 9 (CONT I)

```
                130        140        150        160        170        180
RSW1    ANKA............RHQRHGEEFSSSSRHESQPIPLLTHGHTVSGEIRTPDTQSVRTT
Ath-A   DPEHAAEAALSSRLNTGRGGLDSAPPGSQIPLLTYCDEDADMYSDRHALIVPPSTGYGNR
S0542   NGKGPEWQI.......QRQGEDVDLSSSSRHEQHRIPRLTSGQQISGEIPDASPDRHSIR
Ath-B   K.EKISERMLGWHLTRGKGEEMGEPQYDKEVSHNHLPRLTSRQDTSGEFSAASPERLSVS
Cel-A1  ........................................................
Cel-A2
D48636

                190        200        210        220        230        240
RSW1    SGPLGPSDRNAISSPYIDPRQPVPVRIVDPSKDLNSYGLGNVDWKERVEGWKLKQEKNML
Ath-A   VY..........PAPFTDSSAPPQARSMVPQKDIAEYGYGSVAWKDRMEVWKRRQGEKLQ
S0542   SG..........TSSYVDPSVPVPVRIVDPSKDLNSYGINSVDWQERVASWRNKQDKNMM
Ath-B   ST..........IAGGKRLPYSSDVNQSPNRRIVDPVGLGNVAWKERVDGWKMKQEKNTG
Cel-A1  ....................HARHISSVSTLDSEMAEDNGNSIWKNRVESWKEKKNKKKK
Cel-A2
D48636                                          STTRPGNVAWKERVDGWKLKQDKGAI
```

FIGURE 9 (CONT II)

```
                250        260        270        280        290        300
RSW1     QMT......GKYHEGKGGEIEGTGSNGEELQMADDTRLPMSRVVPIPSSRLTPYRVVIIL
Ath-A    VIK......HEGGNNGRGSNDDDELDDPDMPMMDEGRQPLSRKLPIRSSRINPYRMLILC
S0542    QVA......NKYPEARGGDMEGTGSNGEDIQMVDDARLPLSRIVPIPSNQLNLYRIVIIL
Ath-B    PV....STQAASERGGVDIDASTDILADEALLNDEARQLLSRKVSIPSSRINPYRMVIML
Cel-A1   PAT..........TKVEREAEIPPEQQMEDKPAPDASQPLSTIIPIPKSRLAPYRTVIIM
Cel-A2
D48636   PMTNGTSIAPSEGRGVGDIDASTDYNMEDALLNDETRQPLSRKVPLPSSRINPYRMVIVL

                310        320        330        340        350        360
RSW1     RLIILCFFLQYRTTHPVKNAYPLWLTSVICEIWFAFSWLLDQFPKWYPINRETYLDRLAI
Ath-A    RLAILGLFFHYRILHPVNDAYGLWLTSVICEIWFAVSWILDQFPKWYPIERETYLDRLSL
S0542    RLIILMFFFQYRVTHPVRDAYGLWLVSVICEIWLPLSWLLDQFPKWYPINRETYLDRLAL
Ath-B    RLVILCLFLHYRITNPVPNAFALWLVSVICEIWFALSWILDQFPKWFPVNRETYLDRLAL
Cel-A1   RLIILGLFFHYRVTNPVDSAFGLWLTSVICEIWFAFSWVLDQFPKWYPVNRETYIDRLSA
Cel-A2
D48636   RLVVLSIFLHYRITNPVRNAYPLWLLSVICEIWFALSWILDQFPKWFPINRETYLDRLAL
```

FIGURE 9 (CONT III)

```
                370        380        390        400        410        420
RSW1     RYDRDGEPSQLVPVDVFVSTVDPLKEPPLVTANTVLSILSVDYPVDKVACYVSDDGSAML
Ath-A    RYEKEGKPSGLAPVDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVACYVSDDGAAML
S0542    RYDREGEPSQLAPIDVFVSTVDPLKEPPLITANTVLSILAVDYPVDKVSCYVSDDGSAML
Ath-B    RLVILCLFLHYRITNPVPNAFALWLVSVICEIWFALSWILDQFPKWFPVNRETYLDRLAL
Cel-A1   RYEREGEPDELAAVDFFVSTVDPLKEPPLITANTVLSILALDYPVDKVSCYISDDGAAML
Cel-A2
D48636   RYDREGEPSQLAAVDIFVSTVDPMKEPPLVTANTVLSILAVDYPVDKVSCYVSDDGAAML

                430        440        450        460        470        480
RSW1     TFESLSETAEFAKKWVPFCKKFNIEPRAPEFYFAQKIDYLKDKIQPSFVKERRAMKREYE
Ath-A    TFEALSDTAEFARKWVPFCKKFNIEPRAPEWYFSQKMDYLKNKVHPAFVRERRAMKRDYE
S0542    TFEALSETAEFARKWVPFCKKHNIEPRAPEFYFAQKIDYLKDKIQPSFVKERRAMKREYE
Ath-B    SFESLAETSEFARKWVPFCKKYSIEPRAPEWYFAAKIDYLKDKVQTSFVKDRRAMKREYE
Cel-A1   TFESLVETADFARKWVPFCKKFSIEPRAPEFYFSQKIDYLKDKVQPSFVKERRAMKRDYE
Cel-A2                  RRWVPFCKKHNVEPRAPEFYFNEKIDYLKDKVHPSFVKERRAMKREYE
D48636   TFDALAETSEFARKWVPFVKKYNIEPRAPEWYFSQKIDYLKDKVHPSFVKDRRAMKREYE
```

FIGURE 9 (CONT IV)

```
                490        500        510        520        530        540
RSW1     EFKVRINALVAKAQKIPEEGWTMQDGTPWPGNNTRDHPGMIQVFLGHSGGLDTDGNELPR
Ath-A    EFKVKINALVATAQKVPEEGWTMQDGTPWPGNNVRDHPGMIQVFLGHSGVRDTDGNELPR
S0542    EFKVRINALVAKAQKVPEEGWTMADGTAWPGNNPRDHPGMIQVFLGHSGGLDTDGNELPR
Ath-B    EFKIRINALVSKALKCPEEGWVMQDGTPWPGNNTGDHPGMIQVFLGQNGGLDAEGNELPR
Cel-A1   EYKIRINALVAKAQKTPDEGWTMQDGTSWPGNNPRDHPGMIQVFLGYSGARDIEGNELPR
Cel-A2   EFKVRINALVAKAQKKPEEGWVMQDGTPWPGNNTRDHPGMIQVYLGSAGALDVDGKELPR
D48636   EFKVRINGLVAKAQKVPEEGWIMQDGTPWPGNNTRDHPGMIQVFLGHSGGLDTEGNELPR

                550        560        570        580        590        600
RSW1     LIYVSREKRPGFQHHKKAGAMNALIRVSAVLTNGAYLLNVDCDHYFNNSKAIKEAMCFMM
Ath-A    LVYVSREKRPGFDHHKKAGAMNSLIRVSAVLSNAPYLLNVDCDHYINNSKAIRESMCFMM
S0542    LVYVSREKRPGFQHHKKAGAMNALIRVSAV
Ath-B    LVYVSREKRPGFQHHKKAGAMNALVRVSAVLTNGPFILNLDCDHYINNSKALREAMCFLM
Cel-A1   LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYVNNSKAVREAMCFLM
Cel-A2   LVYVSREKRPGYQHHKKAGAENALVRVSAVLTNAPFILNLDCDHYINNSKAMREAMCFLM
D48636   LVYVSREKRPGFQHHKKAGAMNALVRVSAVLTNGQYMLNLDCDHYINNSKALREAMCFLM
```

FIGURE 9 (CONT V)

```
                610        620        630        640        650        660
RSW1     DPAIGKKCCYVQFPQRFDGIDLHDRYANRNIVFFDINMKGLDGIQGPVYVGTGCCFNRQA
Ath-A    DPQSGKKVCYVQFPQRFDGIDRHDRYSNRNVVFFDINMKGLDGIQGPIYVGTGCVFRKQA
S0542
Ath-B    DPNLGKQVCYVQFPQRFDGIDKNDRYANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTA
Cel-A1   DPQVGRDVCYVQFPQRFDGIDRSDRYANRNTVFFDVNMKGLDGIQGPVYVGTGCVFNRQA
Cel-A2   DPQFGKKLCYVQFPQRFDGIDRHDRYANRNVVFFDINMLGLDGLQGPVYVGTGCVFNRQA
D48636   DPNLGRSVCYVQFPQRFDGIDRNDRYANRNTVFFDINLRGLDGIQGPVYVGTGCVFNRTA

                670        680        690        700        710        720
RSW1     LYGYDPVLTEEDLEPNIIVKSCCGSRKKGKSSKKYNYE.....................KRR
Ath-A    LYGFDAPKKKKPPGKTCNCWPKWCCLCCGLRKKSKTKA.....................KDKKT
S0542
Ath-B    LYGYEPPIKVKHKKPSLLSKLCGGSRKKNSKAKKESDK.....................KKSGR
Cel-A1   LYGYGPPSMPSFPKSSS....SSCSCCCPGKKEPKDPS.....................ELYRDA
Cel-A2   LYGYDPPVSEKRPKMTCDCWPSWCCCCCGGSRKKSKKKGEKKGLLGGLLYGKKKKMMGKN
D48636   LYGYEPPIKQKKKGSFLSSLCGGRKKASKSKKKSSDK.....................KKSNK
```

FIGURE 9 (CONT VI)

```
                730        740        750        760        770        780
RSW1    GINRSDSNAPLFNMEDIDEGFEGYDDERSILMSQRSVEKRFGQSPVFIAATFMEQGGIPP
Ath-A   NTKETSKQIHALENVDEGVIVPVSNVEKRSEATQLKLEKKFGQSPVFVASAVLQNGGVPR
S0542
Ath-B   HTDS.TVPVFNLDDIEEGVEGAGFDDEKALLMSQMSLEKRFGQSAVFVASTLMENGGVPP
Cel-A1  KREELDAAIFNLREIDN.....YDEYERSMLISQTSFEKTFGLSSVFIESTLMENGGVAE
Cel-A2  YVKKGSAPVFDLEEIEEGLEG.YEELEKSTLMSQKNFEKRFGQSPVFIASTLMENGGLPE
D48636  HVDS.AVPVFNLEDIEEGVEGAGFDDEKSLLMSQMSLEKRFGQSAAFVASTLMEYGGVPQ

                790        800        810        820        830        840
RSW1    TTNPATLLKEAIHVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHARGWISIYCNPPR
Ath-A   NASPACLLREAIQVISCGYEDKTEWGKEIGWIYGSVTEDILTGFKMHCHGWRSVYCMPKR
S0542
Ath-B   SATPENFLKEAIHVISCGYEDKSDWGMEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKL
Cel-A1  SANPSTLIKEAIHVISCGYEEKTAWGKEIGWIYGSVTEDILTGFKMHCRGWRSIYCMPLR
Cel-A2  GTNSTSLIKEAIHVISCGYEEKTEWGKEIGWIYGSVTEDILTGFKMHCRGWKSVYCVPKR
D48636  SATPESLLKEAIHVISCGYEDKTEWGTEIGWIYGSVTEDILTGFKMHARGWRSIYCMPKR
```

FIGURE 9 (CONT VII)

```
                850        860        870        880        890        900
RSW1     PAFKGSAPINLSDRLNQVLRWALGSIEILLSRHCPIWYGYHG.RLRLLERIAYINTIVYP
Ath-A    AAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPIWYGYGG.GLKWLERFSYINSVVYP
S0542
Ath-B    PAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGYNG.RLKFLERFAYVNTTIYP
Cel-A1   PAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGFGGGRLKWLQRLAYINTIVYP
Cel-A2   PAFKGSAPINLSDRLHQVLRWALGSVEIFLSRHCPLWYGYGG.KLKWLERLAYINTIVYP
D48636   PAFKGSAPINLSDRLNQVLRWALGSVEILFSRHCPIWYGYGG.RLKFLERFAYINTTIYP

                910        920        930        940        950        960
RSW1     ITSIPLIAYCILPAFCLITDRFIIPEISNYASIWFILLFISIAVTGILELRWSGVSIEDW
Ath-A    WTSLPLIVYCSLPAVCLLTGKFIVPEISNYAGILFMLMFISIAVTGILEMQWGGVGIDDW
S0542
Ath-B    ITSIPLLMYCTLLAVCLFTNQFIIPQISNIASIWFLSLFLSIFATGILEMRWSGVGIDEW
Cel-A1   FTSLPLIAYCSLPAICLLTGKFIIPTLSNLASVLFLGLFLSIIVTAVLELRWSGVSIEDL
Cel-A2   FTSIPLLAYCTIPAVCLLTGKFIIPTLSNLTSVWFLALFLSIIATGVLELRWSGVSIQDW
D48636   LTSIPLLIYCVLPAICLLTGKFIIPEISNFASIWFISLFISIFATGILEMRWSGVGIDEW
```

FIGURE 9 (CONT VIII)

```
                970       980       990      1000      1010      1020
RSW1    WRNEQFWVIGGTSAHLFAVFQGLLKVLAGIDTNFTVTSKATDEDGDFAELYIFKWTALLI
Ath-A   WRNEQFWVIGGASSHLFALFQGLLKVLAGVNTNFTVTSKAAD.DGAFSELYIFKWTTLLI
S0542
Ath-B   WRNEQFWVIGGVSAHLFAVFQGILKVLAGIDTNFTVTSKASDEDGDFAELYLFKWTTLLI
Cel-A1  WRNEQFWVIGGVSAHLFAVFQGFLKMLAGIDTNFTVTAKAAD.DADFGELYIVKWTTLLI
Cel-A2  WRNEQFWVIGGVSAHLFAVFQGLLKVLAGVDTNFTVTAKAAD.DTEFGELYLFKWTTLLI
D48636  WRNEQFWVIGGISAHLFAVFQGLLKVLAGIDTNFTVTSKASDEDGDFAELYMFKWTTLLI

                1030      1040      1050      1060      1070      1080
RSW1    PPTTVLLVNLIGIVAGVSYAVNSGYQSWGPLFGKLFFALWVIAHLYPFLKGLLGRQNRTP
Ath-A   PPTTLLIINIIGVIVGVSDAISNGYDSWGPLFGRLFFALWVIVHLYPFLKGMLGKQDKMP
S0542
Ath-B   PPTTLLIVNLVGVVAGVSYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTP
Cel-A1  PPTTLLIVNMVGVVAGFSDALNKGYEAWGPLFGKVFFSFWVILHLYPFLKGLMGRQNRTP
Cel-A2  PPTTLIILNMVGVVAGVSDAINNGYGSWGPLFGKLFFAFWVILHLYPFLKGLMGRQNRTP
D48636  PPTTILIINLVGVVAGISYAINSGYQSWGPLFGKLFFAFWVIVHLYPFLKGLMGRQNRTP
```

FIGURE 9 (CONT IX)

```
                1090        1100        1110        1120
RSW1    TIVIVWSVLLASIFSLLWVRINPFVDANPNANNFNGKGGVF
Ath-A   TIIVVWSILLASILTLLWVRINPFVAK.GGPVLEICGLNCGN
S0542
Ath-B   TIVVVWSVLLASIFSLLWVRIDPFTSRVTGPDILECGINC
Cel-A1  TIVVLWSVLLASVFSLVWVRINPFVSTADSTTVSQSCISIDC
Cel-A2  TIVVLWSILLASIFSLVWVRIDPFLPKQTGPVLKQCGVEC
D48636  TIVVVWAILLASIFSLLWVRIDPFTTRVTGPDTQTCGINC
```

FIGURE 9 (CONT X)

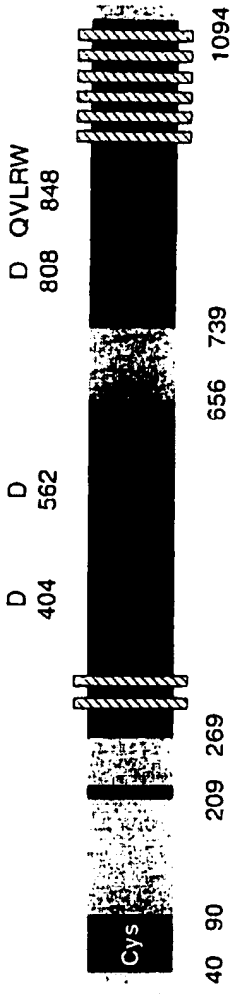


FIGURE 10

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 97/00402

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl[6]: C12N 15/54, 9/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Electronic Database Box below

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Electronic Database Box below

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT. Medline, ChemAbs, Genebank, Swiss Prot, EMBL
Search Terms: Cellulose Biosynthesis, Cellulose Synthase, Sequence ID# 2.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 91/13988 (THE BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM), 19 September 1991 see whole document	1-4, 14-16, 30-32
X	WO 92/18631 (WEYERHAESUR COMPANY) 29 October 1992 see whole document	1-4, 14-16
X	WO 90/12098 (CETUS CORPORATION) 18 October 1990 see whole document	1-4, 14-16

☐ Further documents are listed in the continuation of Box C

☒ See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
14 August 1997

Date of mailing of the international search report
18 AUG 1997

Name and mailing address of the ISA/AU
AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION
PO BOX 200
WODEN ACT 2606
AUSTRALIA Facsimile No.: (06) 285 3929

Authorized officer
Philippa Wyrdeman
for JIM CHAN
Telephone No.: (06) 283 2340

INTERNATIONAL SEARCH REPORT

...ternational Application No.
PCT/AU 97/00402

Box I **Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

Box II **Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Internauonal Application No.
PCT/AU 97/00402

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Repon		Patent Family Member					
WO	9113988	AU	75569/91				
WO	9012098	AU	54373/90	CA	2014264	EP	471687
		IL	94053	NZ	233312	US	5268274
WO	9218631	US	5268274	NZ	233312	CA	2014264
		IL	94053	AU	54373/90	EP	471687

END OF ANNEX